International Journal of Transportation Science and Technology xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

journal homepage: www.elsevier.com/locate/ijtst

Research Paper

A robust deep learning-based system for pedestrian-aware collision prevention in autonomous vehicles

Wajdi Farhat [a,b,*], Marwa Guizani [a,b], Olfa Ben Rhaiem [c], Radhia Zaghdoud [c], Hassene Faiedh [a,b], Chokri Souani [a,b]

[a] Higher Institute of Applied Sciences and Technology, Sousse University, Sousse, Tunisia
[b] Electronics and Microelectronics Laboratory, Faculty of Sciences, University of Monastir, Monastir, Tunisia
[c] Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia

ARTICLE INFO

Article history:
Received 10 April 2025
Received in revised form 4 May 2025
Accepted 29 May 2025
Available online xxxx

Keywords:
Advanced Driver-Assistance Systems (ADAS)
Automatic Emergency Braking (AEB)
Deep learning
Collision avoidance
YOLOv9

ABSTRACT

Pedestrian collisions remain a critical concern in road safety research due to pedestrians' heightened vulnerability and the substantial impact on road networks and public healthcare systems. These incidents often lead to significant traffic disruptions, emphasizing the importance of pedestrian-focused road design. In dynamic urban environments, understanding street layouts and pedestrian behavior is essential for improving safety. In response to these challenges, this research presents an innovative and efficient automated collision avoidance system, leveraging an enhanced YOLOv9 architecture to achieve high accuracy and real-time performance. The proposed system is designed to detect and predict potential pedestrian collisions, while also delivering proactive accident warnings and autonomously initiating the Automatic Emergency Braking (AEB) mechanism when required, thereby effectively mitigating collision risks. By analyzing key behavioral and environmental factors including pedestrian exposure, jaywalking, intersection crowding, and hazardous turning locations, the system employs a proactive methodology aimed at enhancing pedestrian safety. Experimental evaluations using the CityPersons and KITTI datasets demonstrate the system's efficacy, achieving precision rates of 98.79% and 96.62%, respectively, and mAP@0.5 scores of 96.21% and 94.83%. These results underscore the system's capability to deliver high accuracy, fast inference, and proactive accident mitigation across diverse urban scenarios.

1. Introduction

The rapid expansion of global vehicle traffic has raised significant concerns regarding road safety, traffic congestion, environmental sustainability, and public health. Among all vulnerable road users, pedestrians face the highest risk of injury and fatality, especially at urban intersections where traffic complexity is greatest. According to the National Highway Traffic Safety Administration, pedestrian fatalities accounted for a substantial portion of the 42,795 motor vehicle crash deaths recorded in 2022 (Nascimento, 2020; NHTSA, 2022). These incidents not only endanger human life but also contribute to considerable traffic disruptions, emergency response challenges, and broader societal costs.

* Corresponding author at: Higher Institute of Applied Sciences and Technology, Sousse University, Sousse, Tunisia.
E-mail address: wajdifarhat14@gmail.com (W. Farhat).

https://doi.org/10.1016/j.ijtst.2025.05.007
2046-0430/

Please cite this article as: W. Farhat, M. Guizani, O. Ben Rhaiem et al., A robust deep learning-based system for pedestrian-aware collision prevention in autonomous vehicles, International Journal of Transportation Science and Technology, https://doi.org/10.1016/j.ijtst.2025.05.007

To address these issues, the developing proactive, pedestrian-focused safety systems has become a critical necessity. Autonomous Vehicles (AVs), powered by advancements in Artificial Intelligence (AI) and sensor technology, present a promising opportunity to enhance road safety and traffic flow (Arora et al., 2022; Guerrieri and Parla, 2022; Muhammad et al., 2021). However, despite impressive technological achievements, the deploying of AVs in dynamic real-world environments remains challenging, particularly in accurately detecting and predicting pedestrian movements.

Several research efforts have aimed to improve pedestrian detection and collision avoidance. Traditional approaches, relying on motion cues (Bouguettaya et al., 2022; Liang et al., 2022) and hand-crafted features such as HOG and SIFT (Yelchuri et al., 2022; Assefa et al., 2022), established the foundation for early detection systems. The advent of deep learning has significantly improved detection accuracy, with Convolutional Neural Network (CNN)-based models like Faster R-CNN (Dai, 2021), SSD (Yan et al., 2022), and the YOLO family (Hussain, 2024) showcasing remarkable performance. Recent studies have investigated multi-scale feature fusion, attention mechanisms, and temporal modeling to enhance pedestrian detection; however significant challenges remain in detecting small, occluded, or fast-moving pedestrians under diverse urban conditions.

Despite these advancements, existing systems primarily focus on detection accuracy without adequately addressing dynamic risk assessment or timely intervention mechanisms. Specifically, most pedestrian detection frameworks do not proactively generate accident warnings or activate emergency braking based on real-time behavioral and environmental analysis. This gap restricts their practical deployment in highly dynamic and unpredictable urban environments. Therefore, this study proposes a robust, real-time pedestrian-aware collision avoidance system that integrates enhanced pedestrian detection, proactive accident warning, and automatic emergency braking, all within an optimized YOLOv9-based architecture tailored for autonomous vehicles. The main contributions of this paper are as follows:

- We design a real-time pedestrian detection and collision prevention system based on an enhanced YOLOv9 model.
- We integrate advanced modules such as SimAM attention and GSConv to improve small-object detection accuracy.
- We introduce a proactive accident warning and emergency braking mechanism based on real-time risk assessment.
- We validate the system's performance across multiple challenging datasets, demonstrating high precision, recall, and real-time processing capability.

The rest of this paper is organized as follows: Section 2 reviews the existing literature on pedestrian detection and collision avoidance systems. Section 3 describes the proposed methodology. Section 4 presents experimental evaluations and results, while Section 5 concludes the study with future research directions.

2. Literature review

This section presents a concise overview of pedestrian detection within the context of intelligent transportation systems and autonomous vehicle technologies. Vision-based pedestrian detection algorithms are generally classified into three categories: motion-based approaches, hand-crafted feature-based methods, and deep learning-based techniques.

Motion-based approaches include frame subtraction, optical flow, and background subtraction methods (Chen, 2021; Chaurasiya and Ganotra, 2023). Frame subtraction detects moving objects by calculating pixel-wise differences between consecutive frames. While this method is computationally efficient and adaptable to dynamic backgrounds, it often fails to capture extremely fast or slow movements accurately. Optical flow techniques estimate the motion vectors of individual pixels over time, offering high accuracy but at the expense of significant computational complexity and processing time (Sormoli et al., 2024). Background subtraction methods, such as Gaussian Mixture Models (GMM), are widely employed to differentiate foreground from background elements. However, their performance deteriorates in scenarios involving stationary pedestrians (Ha et al., 2020).

Prior to the rise of deep learning, pedestrian detection predominantly relied on hand-crafted features. Methods such as Scale-Invariant Feature Transform (SIFT) (Costea et al., 2017); Haar-like features, and Histograms of Oriented Gradients (HOG) (Asha et al., 2022) contributed significantly to early progress in the field. Notably, the Deformable Part-Based Model (DPM) (Zhang et al., 2024) achieved improved accuracy by leveraging refined HOG features to represent individual pedestrian parts with greater detail. These approaches were typically coupled with classifiers such as Support Vector Machines (SVM) and AdaBoost. Although effective in specific contexts, hand-crafted feature-based methods are inherently limited in their ability to capture complex visual patterns, making them less competitive than contemporary deep learning-based models.

Recent advancements in deep learning have revolutionized pedestrian detection, with Convolutional Neural Networks (CNNs) demonstrating superior performance in handling complex visual scenarios. Models such as Faster R-CNN (Chen, 2021) and the Single Shot Multibox Detector (SSD) (Yan et al., 2022) leverage region-based and single-stage detection frameworks, respectively, to achieve high accuracy and speed. The YOLO (You Only Look Once) family of models, particularly YOLOv8, has further enhanced real-time capabilities through advanced feature extraction and multi-scale detection mechanisms (Hsu and Lin, 2021). One promising approach to understanding street behaviors in urban environments is through video analytics. Researchers have employed fixed cameras in cities to analyze pedestrian crossing behaviors (Zhang et al., 2020), vehicle–pedestrian interactions (Liang et al., 2021), and pedestrian gap acceptance (Gorrini et al., 2018). Advances in deep learning-based computer vision algorithms have significantly improved the use of video data for capturing pedes-

trian behaviors. Object detection models like Fast R-CNN (Han et al., 2019) and Mask R-CNN (Fang et al., 2023), combined with tracking algorithms such as Deep Sort (Brunetti et al., 2018), can detect pedestrians, predict their behavior, and track their movements.

Pedestrian detection remains a critical component in applications such as autonomous driving and video surveillance. Over the years, a wide range of detection techniques have been developed, each presenting unique strengths and limitations. Table 1 offers a comparative summary of prominent methods, detailing input types, algorithmic strategies, and performance metrics to assist researchers in selecting suitable approaches.

Recent advancements have extended beyond traditional deep learning models to improve detection robustness, processing speed, and environmental awareness key factors for proactive collision avoidance. For instance, YOLOv11 integrates transformer-based feature extraction, significantly boosting detection efficiency in dynamic environments (Wisesa et al., 2025). YOLOv12 further refines this by adopting dynamic label assignment and improved multi-scale feature fusion, enhancing performance in complex urban settings (He et al., 2025). Additionally, semantic scene completion methods like SSCFormer (Wang et al., 2025) reconstruct 3D scenes from partial inputs, enabling systems to better manage occlusions and incomplete visual data. These innovations reflect the field's shift toward holistic environmental perception, essential for reliable safety interventions. Complementing detection improvements, recent studies emphasize modeling pedestrian behavior to enhance decision-making in autonomous vehicles (AVs). Nafakh et al. (Nafakh, 2023) proposed a quantitative framework that translates pedestrian-driver interaction data into adaptive signal timing strategies, improving accessibility for vulnerable users in both conventional and connected autonomous vehicle (CAV) systems. Similarly, John et al. (Labi et al., 2024) introduced a virtual and augmented reality-based simulation framework to study pedestrian-AV interactions under realistic urban scenarios, providing deeper behavioral insights.

Recent studies in the domain of intelligent navigation and behavior prediction have contributed to techniques that can inform pedestrian-aware planning strategies. For instance, reinforcement learning methods such as the A*-enhanced double deep Q-network have shown promise in dynamic route planning under uncertain environments (Chen, 2025). Similarly, orientation-aware object detection models (Chen et al., 2023) have demonstrated the value of rotation feature decoupling for robust spatial interpretation, which aligns with our use of adaptive attention mechanisms. Furthermore, driver behavior modeling studies have used real-world trajectory data to analyze lane-change impacts on following vehicles, offering valuable safety implications for cooperative vehicle systems (Gu et al., 2023).

Understanding driver avoidance behavior is equally crucial for advancing traffic safety. Effective modeling of such behavior under critical conditions supports the development of intelligent driving systems and informed traffic management strategies (Dong et al., 2022; Ji et al., 2024).Avoidance behaviors are generally classified into braking (longitudinal), steering (lateral), or combined maneuvers (Gao et al., 2024; Losada et al., 2023). Although pivotal for collision prevention, most research isolates braking and steering, overlooking their complex interdependence (Park et al., 2021; Zhang and Berger, 2023). Furthermore, the nuances of pedestrian-vehicle interaction often remain underexplored, limiting comprehensive behavioral modeling (Sarkar et al., 2021; Rezwana and Lownes, 2024). The integration of unmanned aerial vehicles (UAVs) into traffic monitoring has introduced new opportunities for collecting vehicle trajectory data, enabling detailed analysis of traffic conflicts (Kušić et al., 2023). Historically, video-based analysis and traffic safety modeling have evolved separately,

Table 1
Comparative Analysis of Pedestrian Detection Methodologies.

Reference	Technique	Major Contribution
(Wang et al., 2023)	YOLO	To identify 2D object bounding boxes in the color image, DM, and RM, three YOLO-based object detectors were independently applied to each mode. The detection results were then merged using an evaluation function and non-maximum suppression techniques.
(Bin Zuraimi and Kamaru Zaman, 2021)	YOLOv4, DeepSORT	DeepSORT utilizes deep learning algorithms to minimize identity shifts and improve tracking performance, particularly during occlusions in the SORT algorithm. This vehicle detection and tracking approach integrates the YOLOv4 model-based DeepSORT algorithm implemented using the TensorFlow library.
(Bouguettaya et al., 2022)	CNN	The early layers of Convolutional Neural Networks (CNNs) extract fundamental low-level vehicle features, while the deeper layers capture more intricate intermediate and high-level representations required for vehicle classification. These networks' performance is refined using learnable parameters such as weights and biases.
(Han et al., 2019)	YOLO, R-CNN	The goal is to compare the training accuracy and performance of YOLOv4 and Faster R-CNN. Datasets comprising images captured under diverse weather conditions, terrains, and times of day have been prepared for use during the training and testing phases.
(Eskandari Torbaghan et al., 2022)	YOLOv3, SSD, R-CNN	Fast and high-performance object detection algorithms, such as Faster R-CNN and YOLOv3, are employed to analyze images captured by UAVs, while lightweight algorithms like Tiny-YOLO and SSD are used to detect vehicles. These methods automatically determine the positions of moving vehicles in traffic.
(Qiu et al., 2021)	FE-CNN, TensorFlow Deep learning framework	By optimizing the number of convolution kernels, the weights and biases in the neural network are maximized during CNN training. Consequently, FE-CNN achieves recognition precision comparable to GoogleLeNet in less time, with stable precision that surpasses GoogleLeNet's performance. TensorFlow, leveraging the generated computation graph, identifies all nodes dependent on E.

with limited interdisciplinary efforts (Shawky et al., 2023). Merging these domains offers deeper insights into avoidance behavior, supporting the development of intelligent traffic systems and improved safety interventions (Fang et al., 2024).

Despite notable progress, the application of these techniques to domain-specific contexts such as bus transit safety remains limited. Expanding research in this area could yield valuable insights into pedestrian dynamics and support targeted safety enhancements within public transportation networks.

3. Methodology

This section presents the proposed automated collision avoidance system, developed to improve vehicle safety. Leveraging AI-powered cameras, the system continuously monitors the vehicle's surroundings, identifying and responding to potential threats in real time. It processes video data to detect pedestrians and obstacles, evaluates their type, location, and distance, and assesses collision risks. By utilizing advanced deep learning algorithms, the system ensures accurate risk detection and prompt intervention to help prevent accidents. The system integrates a particle filter to temporally smooth pedestrian trajectories and enhance Time-to-Collision estimation under uncertainty. The system's key components include:

- **Pedestrian Detection Module**: This module performs real-time detection and classification of pedestrians, offering precise location data to enable accurate tracking and monitoring.
- **Collision Risk Assessment Module**: This module evaluates whether a detected pedestrian poses a collision risk based on their position relative to the vehicle's trajectory. It issues appropriate warnings or activates emergency braking mechanisms as needed, ensuring timely responses to mitigate potential accidents.

The pedestrian detection system, illustrated in Fig. 1, identifies pedestrians in real time and evaluates their distance from the vehicle. Using this data, the system calculates the required breakdown to prevent potential collisions. The figure highlights how distance intervals are employed to trigger emergency braking in urban environments, ensuring timely and accurate responses to potential hazards.

A major challenge in autonomous driving is managing threats posed by nearby objects, such as pedestrians. Accidents often occur when vehicles fail to stop in time as pedestrians cross their path. Preventing such incidents requires early threat detection and precise braking actions. However, variables such as vehicle speed, pedestrian position and velocity, crossing timing, direction, sensor inaccuracies, and road conditions introduce significant uncertainty into braking system design. Even with accurate pedestrian detection, determining when a pedestrian constitutes a collision threat is challenging. To address this, the braking system must evaluate the pedestrian's state, including their position and velocity, and decide on appropriate actions accordingly.

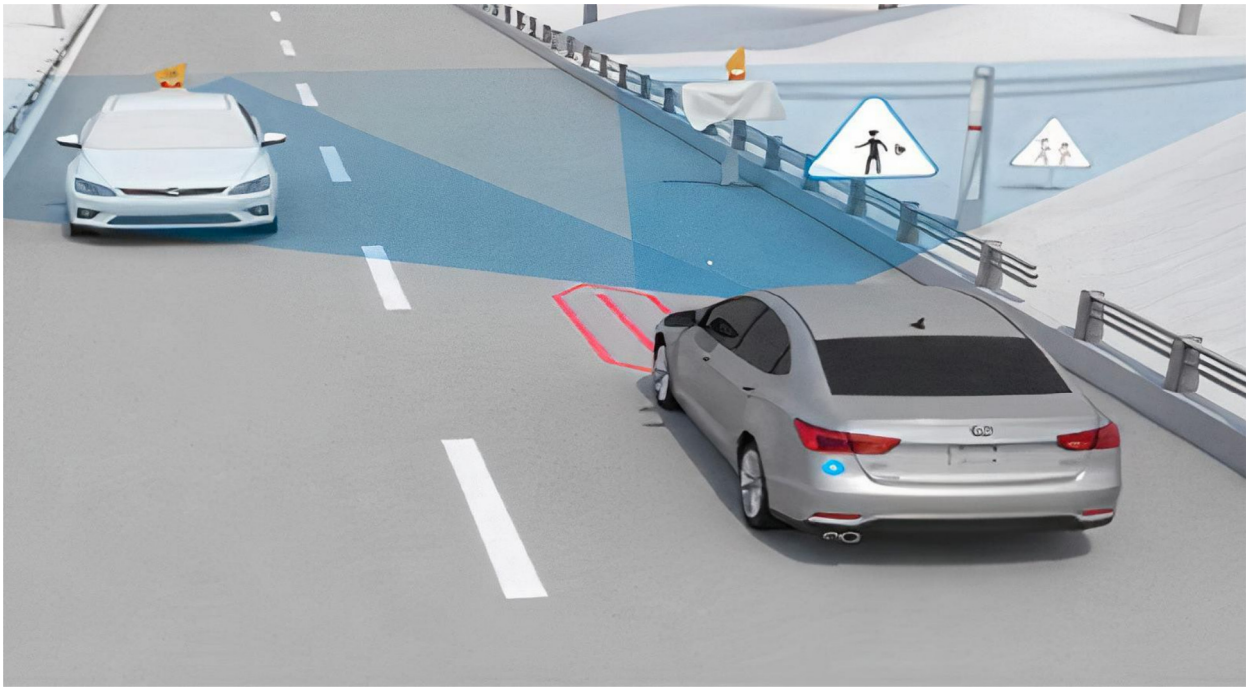

Fig. 1. Overview of the Automated Collision Avoidance System.

3.1. Pedestrian detection architecture

3.1.1. Pedestrian region identification

Accurate identification of pedestrian regions is essential for effective pedestrian detection systems, enabling precise boundary delineation and facilitating subsequent tasks such as action recognition or collision avoidance. Traditional image processing methods often struggle with challenges like variable lighting, occlusions, and diverse backgrounds, which limit their reliability in real-world applications. To overcome these challenges, we employ deep learning-based semantic segmentation, which has demonstrated significant advances in extracting detailed image features, making it highly effective for pedestrian detection.

Our proposed architecture builds upon semantic segmentation techniques but introduces key innovations. The network consists of an encoding section and a decoding section. The encoding section leverages convolution and pooling layers to extract hierarchical features and capture both local and global contextual information. To enhance the network's ability to detect pedestrians in dynamic environments, we incorporate dilated convolution layers, which allow for multi-scale feature extraction without increasing computational overhead. These layers expand the receptive field, enabling the model to capture a wider range of contextual information.

The decoding section of the network focuses on restoring the spatial resolution of the feature maps, facilitating precise pedestrian region segmentation. This section uses up sampling and deconvolution layers to recover the original image resolution while maintaining the important semantic features. Finally, the network employs a SoftMax layer for pixel-wise classification, assigning each pixel to a specific class, such as pedestrian or background, ensuring accurate segmentation.

This architecture, illustrated in Fig. 2, integrates advanced techniques to address the complexities of pedestrian detection in challenging environments. By combining hierarchical feature extraction, multi-scale representation, and efficient resolution restoration, our approach significantly enhances detection accuracy, even under occlusions and varying lighting conditions. To improve the system's robustness in dynamic environments and account for uncertainty in pedestrian motion, a particle filter-based temporal modeling component was integrated into the post-detection stage of the framework. The objective is to produce smoother, more stable pedestrian trajectories and enhance the reliability of Time-to-Collision (TTC) estimation, particularly under challenging conditions such as partial occlusion, abrupt movement, and noisy sensor data.

The particle filter operates as a recursive Bayesian estimator, modeling each pedestrian's state (typically position and velocity) using a set of NNN weighted particles. Each particle represents a hypothesis of the pedestrian's true state. During prediction, particles are propagated forward using a constant velocity motion model with added process noise to account for variability. Upon receiving a new observation from the detection module, the filter updates the particle weights using a Gaussian likelihood function comparing the detected position with each particle's predicted state. At each time step, the weighted particle set is resampled to focus on high-probability regions of the state space, thus refining the position estimate. This temporal smoothing process allows the system to tolerate brief detection dropouts and reduces frame-to-frame jitters in bounding boxes. Most importantly, it improves TTC estimates by providing continuous, noise-filtered velocity and position information, which is critical for reliable and early warning activation. The particle filter is implemented efficiently with low computational overhead and runs in real time, making it suitable for integration with deep learning-based perception systems in autonomous driving pipelines.

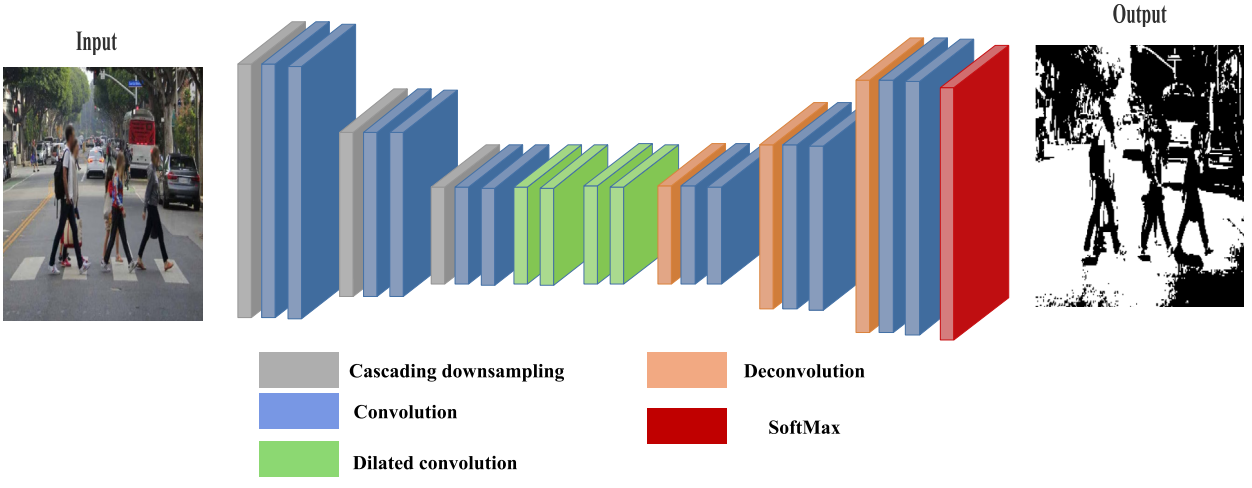

Fig. 2. In-depth Network Architecture for Pedestrian Recognition.

3.1.2. Deep learning-based pedestrian detection

This research presents an advanced approach to pedestrian detection utilizing the YOLOv9-optimized model, chosen for its exceptional accuracy, streamlined design, and real-time processing capabilities. The YOLOv9-optimized architecture is particularly well-suited for safety–critical applications, such as collision avoidance, due to its enhanced feature extraction capabilities and precise object localization. Fig. 3 depicts the YOLOv9-optimized architecture, highlighting its specialized design tailored to achieve accurate and efficient pedestrian detection in diverse environments.

The YOLOv9-Optimized network architecture uses YOLOv9 as the base model, which is further enhanced with a small-object detection head, GSConv convolutional blocks, and the SimAm attention mechanism to significantly improve performance. The refined network structure consists of three primary components: the backbone, the neck, and the detection head.

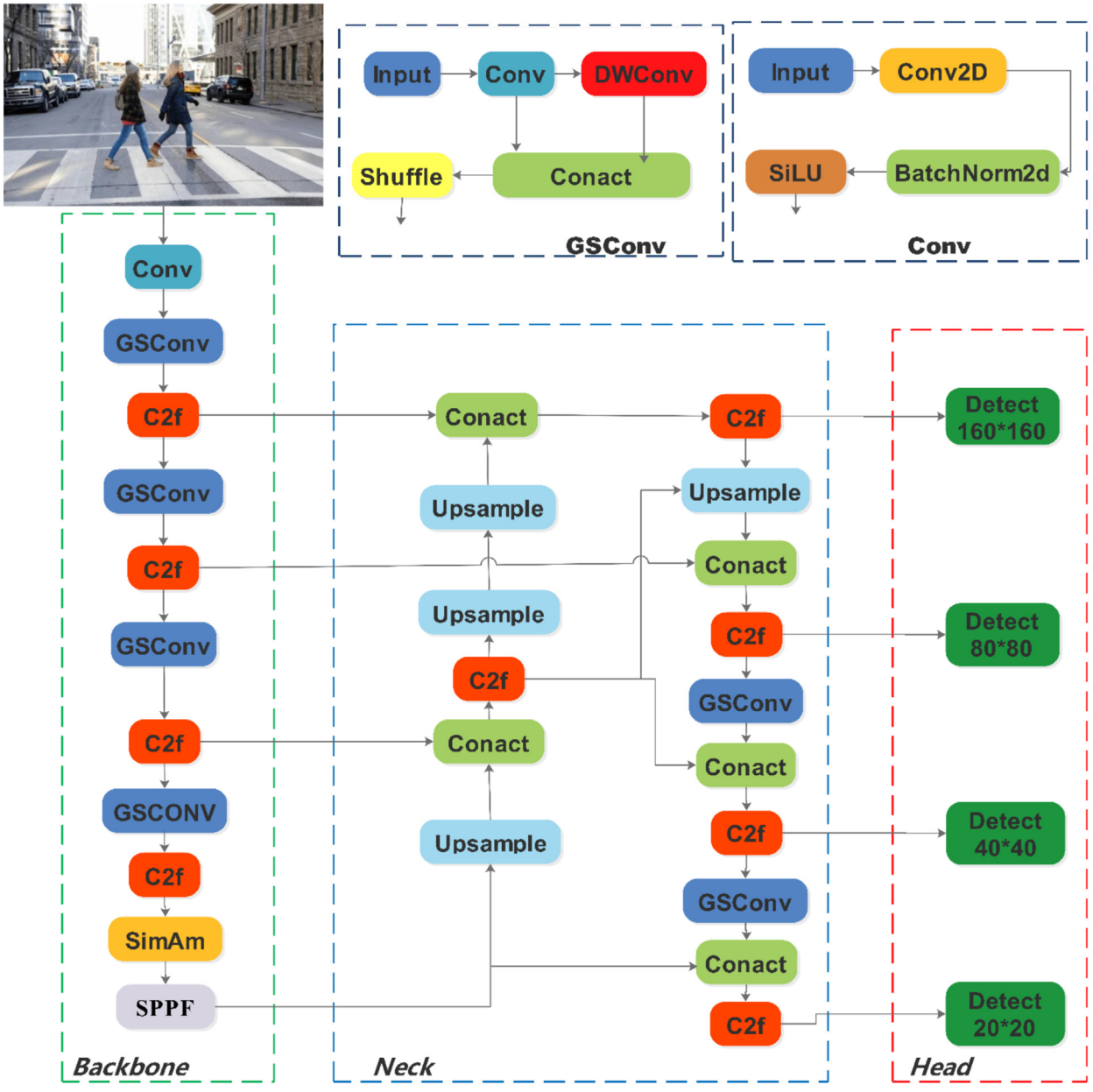

Fig. 3. Architecture of the YOLOv9 Model for Pedestrian Detection.

- **Backbone:** In our architecture, GSConv modules replace standard convolutional blocks in the Neck and Head components to optimize feature fusion and detection efficiency. This includes the detection head branches and fusion layers that benefit from lightweight spatial-channel interaction. The initial convolutional layers in the Backbone particularly the first stage responsible for early feature extraction are preserved as standard Conv modules to maintain spatial fidelity. This selective replacement balances performance with computational efficiency.
- **Attention Mechanism:** The SimAm attention mechanism was incorporated before the Spatial Pyramid Pooling-Fast (SPPF) module. This mechanism effectively suppresses background noise while maintaining parameter efficiency, leading to enhanced detection performance.
- **Neck:** To better handle multi-scale detection, we introduced additional pathways and a small-object detection head in the neck section. These modifications improve the model's ability to detect pedestrians and other objects at varying scales, minimizing the risk of overlooking or misclassifying distant pedestrians.
- **Head:** The detection head features a decoupled structure that separates classification and prediction tasks. By processing feature maps at different scales, this structure ensures precise localization and categorization of objects, allowing for accurate detection of pedestrians, even in complex and dynamic environments.

YOLOv9 models are available in five variants, differentiated by network depth, width, and the maximum number of channels: n, s, m, l, and x. Among these variants, YOLOv9 stands out for its minimal parameters and floating-point operations, making it an efficient choice for real-time applications without sacrificing detection accuracy.

The YOLOv9 architecture includes three key components:

- **Backbone:** Conv modules, C2f modules, and the SPPF module form the backbone, which extracts features from input images. The Conv modules reduce the image resolution and increase the number of channels to produce refined feature maps, while the C2f modules are adept at extracting deep, multi-scale features, thereby improving the model's capacity to recognize complex patterns.
- **Neck:** The neck integrates features from multiple scales using a combination of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). Through operations like upsampling, channel concatenation, and deep feature extraction, the neck ensures effective merging of features from different scales, improving detection accuracy.
- **Head:** The head utilizes a decoupled structure to separate classification and prediction tasks. By operating on feature maps from various scales, this design allows for precise object detection in terms of both size and position, ensuring accurate localization and categorization of objects.

a. Enhanced Detection Head for Small Objects

In the KITTI and City Persons datasets, distant pedestrians occupy only a small portion of the image. When these images are resized to 640×640 pixels, they contain numerous small targets. However, during the subsequent upsampling and pooling operations in the neck network, many features associated with these small targets are lost, leading to missed detections. The baseline model employs detection heads with sizes of 80×80, 40×40, and 20×20. The 80×80 detection head, due to its large receptive field, is less effective for detecting small targets, making it challenging for the baseline model to accurately identify abnormal behavior in distant pedestrians.

To address this limitation, we enhanced the baseline model by introducing a 160×160 detection head specifically designed for small targets. The structural diagram of the modified detection head, with the minimal changes to the original model clearly highlighted for reference. The feature map, which is initially upsampled to 40×40 in the neck layer, undergoes two additional upsampling operations to generate a more detailed 160×160 feature map. This high-resolution feature map captures significantly more information about small targets. To further enrich the feature representation, the 160×160 feature map is concatenated with the corresponding feature map from the second layer of the backbone network. This integration enhances the model's ability to detect small target behaviors at this scale.

To further improve the detection of small-scale pedestrian instances, we extend the detection head design by integrating a multi-scale feature fusion strategy guided by Feature Pyramid Network (FPN) theory. While conventional FPNs perform multi-scale aggregation through upsampling and summation, such approaches may inadequately balance semantic abstraction from deeper layers with the spatial precision of shallower layers, particularly in complex or distant scenes. To address this, we propose an adaptive fusion mechanism that employs learnable weights to modulate the contribution of each scale based on task-specific relevance. Specifically, the fused feature map F_f is computed as:

$$F_f = \sum_{i=1}^{N} \alpha_i \cdot F_i, \text{ with } \sum_{i=1}^{N} \alpha_i = 1, \alpha_i \geq 0 \tag{1}$$

where F_i denotes the feature map at scale i, and α_i is the corresponding learned fusion weight, normalized via a softmax function to ensure interpretability and stability during training.

This adaptive mechanism allows the model to emphasize high-resolution features from the 160×160 head when detecting small or partially visible objects while still integrating semantic cues from coarser levels. By leveraging this formulation, the system can more effectively capture pedestrians at a distance or under occlusion, which often appear as small-scale objects in the image.

b. Modeling Context-Specific Risk Factors

To advance beyond generic pedestrian detection, the proposed system incorporates a context-aware risk modeling framework that implicitly identifies critical behavioral and environmental factors influencing collision likelihood. This is accomplished through a combination of spatial–temporal trajectory tracking, visual scene context analysis, and dynamic adaptation of collision metrics such as TTC. Specifically, jaywalking behavior is identified by detecting trajectory deviations from expected pedestrian movement patterns. Instances in which pedestrians enter the vehicle's path from non-designated crossing zones or move at irregular lateral angles are flagged through particle filter-based trajectory smoothing. These irregular motion patterns lead to a context-sensitive reduction in TTC thresholds, enabling earlier risk warnings and proactive braking activation.

Intersection crowding is addressed by analyzing the spatial density and convergence of detected pedestrians in known intersection regions. Bounding box clustering and overlap metrics are used to quantify pedestrian density. When high-density pedestrian activity is detected near intersection entry points, the system elevates the collision risk level, issuing earlier warnings to compensate for the increased unpredictability in such scenarios. Environmental features such as sharp road curves and intersection geometry are inferred through visual cues, including bounding box alignment distortion, nonlinear motion trajectories, and lateral field-of-view compression in the camera stream. These indicators suggest reduced visibility or blind-spot vulnerability, prompting the system to modify TTC calculations accordingly to ensure timely intervention in constrained navigation conditions.

Pedestrian exposure is quantified using a dwell-time-based occupancy model. By tracking how long a pedestrian remains within a critical risk zone defined as the projected path of the vehicle, the system dynamically adjusts the risk weighting of each tracked individual. Prolonged presence or low-velocity dwell near the vehicle path increases alert sensitivity and escalates response priority. Together, these mechanisms equip the system with a robust understanding of real-world behavioral and environmental dynamics, enhancing its ability to perform proactive collision prevention in complex urban driving environments.

c. GSConv Module

Lightweight network design has increasingly adopted Depth-wise Separable Convolution (DSConv) to reduce model parameters and floating-point operations (Alalwan et al., 2021). DSConv independently performs convolutions across the input channels, effectively minimizing redundant feature information. However, this separation of channel information in DSConv reduces the model's feature extraction capability compared to dense channel convolutions, such as standard convolutions.

As illustrated in Fig. 4, GSConv addresses this limitation by combining the strengths of standard convolution and DSConv. The process begins with a standard convolution applied to the input feature map with $C1$ channels, generating an intermediate feature map with $C_2/2$ channels. Next, DSConv is applied to the intermediate feature map, yielding a second intermediate feature map with $C_2/2$ channels. The two intermediate feature maps are then combined and shuffled to produce the final output feature map with C_2 channels. This approach effectively preserves critical feature information that might otherwise be lost due to the channel separation in DSConv, while maintaining an output comparable to that of standard convolution.

The time complexity formulas for standard convolution (SC), Depth-wise Separable Convolution (DSC), and GSConv are presented in Equations (2), (3), and (4), respectively.

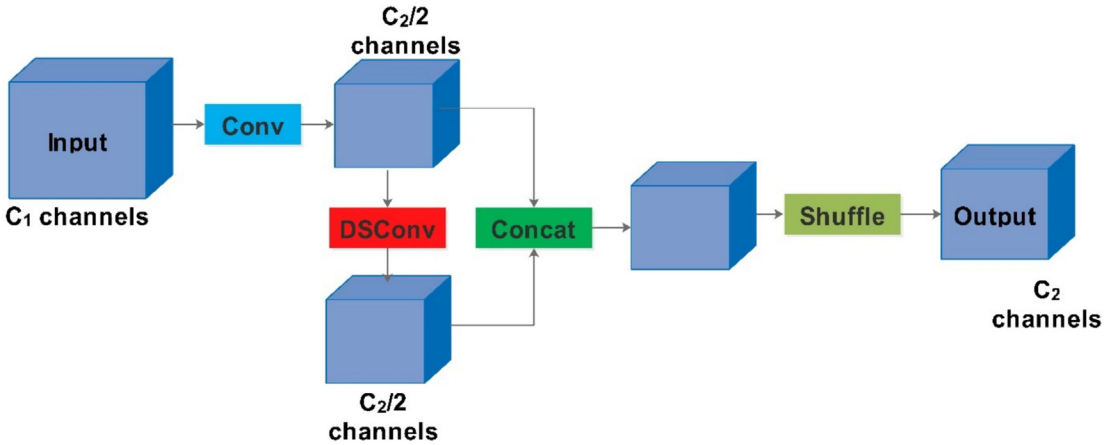

Fig. 4. Framework of the GSConv Module.

$$O(W \cdot H \cdot K_1 \cdot K_2 \cdot C_1 \cdot C_2) \tag{2}$$

$$O(W \cdot H \cdot K_1 \cdot K_2 \cdot C_1) \tag{3}$$

$$O([W \cdot H \cdot K_1 \cdot K_2 \cdot C_1]/2 \cdot (C_1 + 1)) \tag{4}$$

In the above formulas, W and H denote the width and height of the output feature map, K_1 and K_2 refer to the sizes of the convolution kernels, C_1 represents the number of input feature channels, and C_2 indicates the number of output feature channels.

d. SimAm Attention Module

This paper presents SimAm (Qu et al., 2023), a novel parameter-free attention mechanism grounded in neural network theory. To enhance object recognition and mitigate interference from complex backgrounds, the SimAm attention mechanism was strategically positioned before the detection head. This placement effectively reduced noise and improved focus on small objects, such as corn silk, significantly minimizing sea surface interference while greatly enhancing detection performance.

The experimental results underscore the SimAm module's efficacy in suppressing background noise and improving object recognition accuracy. Unlike traditional attention mechanisms that primarily target either the channel or spatial domains, SimAm stands out by seamlessly integrating spatial, channel, and feature dimensions to compute 3D attention weights. The process for generating these 3D attention weights is illustrated in Fig. 5.

In visual neuroscience, active neurons could suppress the activity of neighboring neurons. SimAm leverages this principle by prioritizing neurons that exhibit strong spatial suppression. This prioritization is determined by using an energy function, as described in Eq. (5).

$$e_t(w_t, b_t, x_i) = \frac{1}{M-1}\sum_{i=1}^{M-1}(-1 - (w_t x_i + b_t))^2 + (1 - (w_t + b_t))^2 + \lambda w_t^2 \tag{5}$$

In the formula, t represents the target neuron, while x_i symbolizes the other neurons within a single channel of input features. The spatial dimension index is denoted by i, while $M = H * W$ represents the total number of neurons in the channel. The weight and bias of the target neuron, denoted as w_t and b_t, are determined using the formulas in Eqs. (6) and (7).

$$w_t = -\frac{2(t - \mu_t)}{(t + \mu_t)^2 + 2\sigma_t^2 + 2\lambda} \tag{6}$$

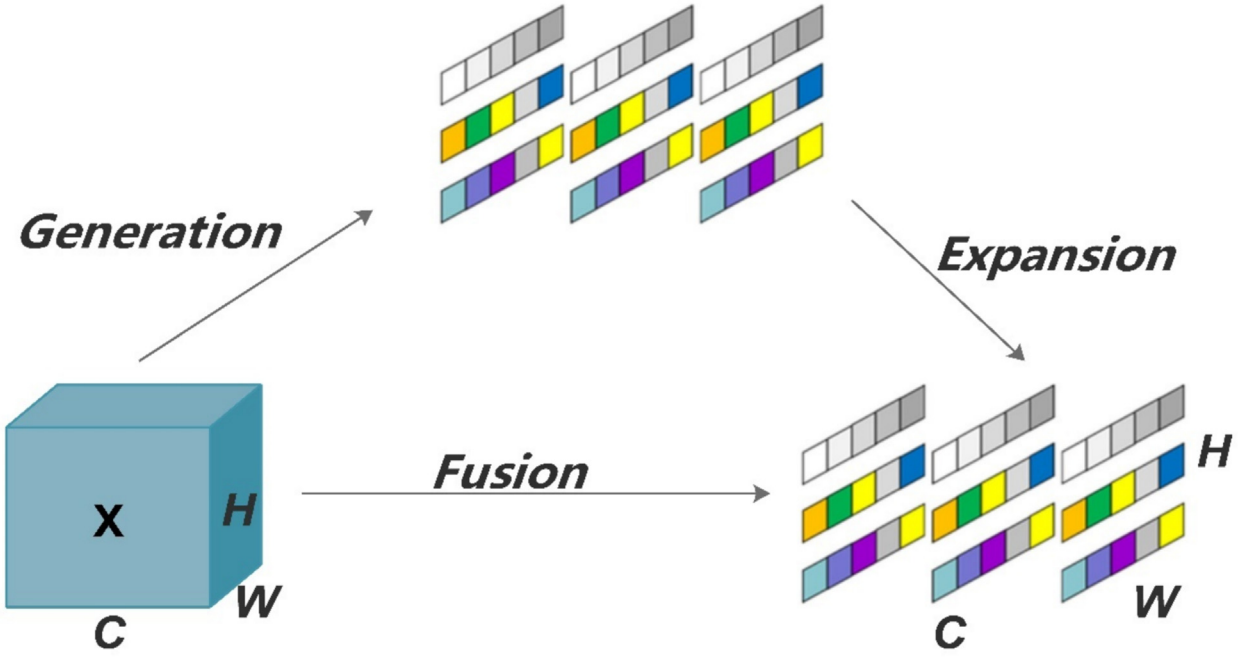

Fig. 5. Schematic diagram of SimAM attention module.

$$b_t = -\frac{1}{2}(t + \mu_t)w_t \tag{7}$$

In the formula, $\frac{1}{M-1}\sum_{i=1}^{M-1} x_i$ and $\sigma_t^2 = \frac{1}{M-1}\sum_{i=1}^{M-1}(x_i - \mu_t)^2$. To ensure that all pixels within a single channel adhere to the same distribution, the minimum energy function can be derived.

$$e_t^* = \frac{4(\widehat{\sigma} + \lambda)}{(t - \widehat{\mu}) + 2\widehat{\sigma}^2 + 2\lambda} \tag{8}$$

In Formula (8), $\widehat{\sigma}$ and $\widehat{\mu}$ represent the mean and variance of all neurons excluding t. A smaller e_t^* value denotes higher importance, with importance calculated as $\frac{1}{e_t^*}$. By refining the energy distribution across neurons, the scaled energy values are used to enhance feature refinement, resulting in the final feature map. This process is represented in Formula (9):

$$\widetilde{X} = sigmoid\left(\frac{1}{E}\right) \odot X \tag{9}$$

Here, $\widetilde{X}$ denotes the refined feature map, sigmoid applies a nonlinear scaling, and $\odot$ represents element-wise multiplication.

The SimAM attention module enhances spatial feature discrimination by assigning importance weights to individual neurons based on an energy-based evaluation of their contribution to the output decision (Tishby et al., 2000). As a parameter-free mechanism, SimAM provides an efficient means of focusing on discriminative spatial features without increasing model complexity, making it particularly well-suited for real-time systems.

From an information-theoretic perspective, SimAM implicitly aligns with the principles of the Information Bottleneck (IB) theory (Yang et al., 2021), which seeks to extract the most relevant information for a target task while discarding irrelevant or redundant input. By suppressing spatially redundant activations and emphasizing task-relevant features, SimAM effectively performs a lightweight form of feature compression. This is especially beneficial in pedestrian detection tasks, where background clutter, lighting variations, and motion blur often introduce noise that impairs model performance.

Although a formal re-derivation of SimAM under the IB framework through mutual information estimation and constrained optimization is beyond the scope of this work, its current design reflects key IB principles. Specifically, SimAM's scoring mechanism approximates an importance-driven compression process that improves attention-guided learning, reduces activation noise, and enhances generalization under dynamic environmental conditions.

3.2. Collision risk assessment

The Autonomous Emergency Braking (AEB) early warning system employs a hierarchical algorithm to classify safety levels based on Time-to-Collision (TTC) values, ensuring timely intervention in pedestrian-related scenarios. To support proactive collision avoidance, three critical safety distances are defined:

- **Warning Safety Distance (WSD):** The minimum threshold at which the system issues an early auditory or visual warning, signaling a potential collision risk.
- **Braking Initiation Safety Distance (BISD):** The point at which the system initiates deceleration to avoid a collision if the pedestrian continues into the vehicle's path.
- **Minimum Braking Safety Distance (MBSD):** The absolute minimum stopping distance required to prevent a collision, assuming maximum braking force.

For instance, consider an autonomous vehicle traveling at 50 km/h (13.9 m/s) approaching a pedestrian crosswalk. The system may issue a warning at 40 m (WSD), initiate braking at 25 m (BISD), and require a full stop within 10 m (MBSD) to ensure pedestrian safety. These thresholds function sequentially to identify risk, initiate appropriate actions, and guarantee safe deceleration under real-world conditions.

The AEB system further categorizes driving scenarios into three safety levels, each associated with distinct system responses and signal values:

1. **Green Zone – Driving Safety Level (Signal = 0):** Indicates normal driving conditions with no immediate threats. The system remains passive.
2. **Orange Zone – Collision Warning Level (Signal = 1):** Represents potential collision risk. The system issues warnings, prompting driver intervention.
3. **Red Zone – Collision Danger Level (Signal = 2):** Signals imminent collision. If the driver does not respond, the system automatically engages the brakes to avoid impact.

These zones, defined by TTC-based thresholds, guide the AEB system's transition from monitoring to active intervention. The system remains inactive in safe conditions, issues continuous alerts under moderate risk, and autonomously intervenes under high-risk scenarios. Fig. 6 illustrates this safety-level classification and the corresponding operational responses.

The accident warning system uses Time-to-Collision (TTC) as the core predictive metric, defined as:

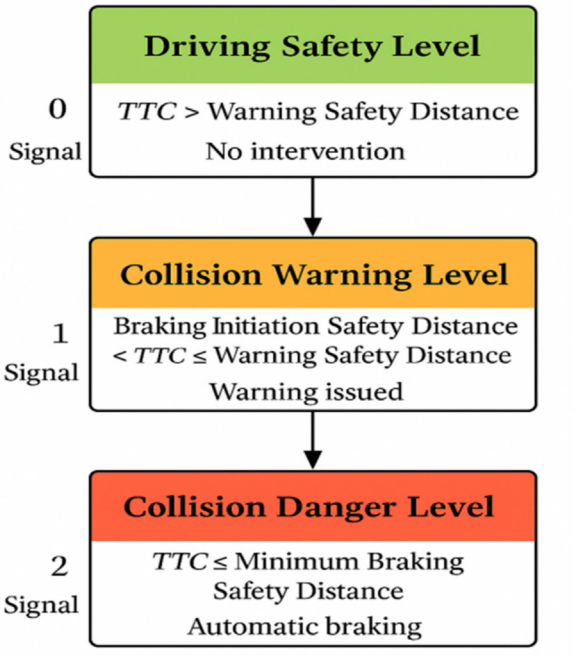

Fig. 6. Classification of Safety Levels in AEB System.

$$TTC = \frac{d}{v}$$

where d is the estimated distance between the vehicle and the pedestrian, and v is the current vehicle velocity. A warning is issued if:

$$TTC < \tau$$

The threshold τ represents the minimum time margin required to ensure that the AEB system can activate and bring the vehicle to a stop before impact. The value of τ is not chosen arbitrarily; it is derived using a cost-sensitive risk-balancing approach grounded in Bayesian decision theory and risk-minimization criteria.

In this framework, the probability of collision is dynamically updated using Bayesian inference, incorporating prior knowledge about vehicle speed, pedestrian movement, sensor accuracy, and environmental conditions. The threshold τ is set to minimize the expected risk of collision, balancing the costs of false negatives (missed warnings) and false positives (unnecessary warnings). Specifically, false negatives are associated with a substantially higher cost than false positives, underscoring the critical importance of prioritizing safety in decision-making outcomes. Through iterative simulation and testing across urban scenarios, the optimal value of τ was empirically validated. This approach ensures high early-warning reliability while minimizing false alarms, making the system suitable for deployment in unpredictable, dynamic pedestrian environments.

The AEB system's three hierarchical warning levels require a corresponding division of the TTC intervals. Among these, determining the appropriate warning time for the second level, which represents the collision warning stage, is critical. This parameter significantly impacts the system's reliability and its effectiveness in preventing collisions. If the warning time is too short, drivers may not have enough time to respond, reducing the system's ability to effectively warn against pedestrian collisions. On the other hand, if the warning time is excessively long, it may disrupt normal driving, leading to driver frustration and diminished trust in the AEB system. Therefore, setting the optimal duration for the second-level warning is a key consideration. To achieve this, factors such as driver reaction time, the time required to close brake gaps, and hydraulic lag must be carefully evaluated. Additionally, a thorough analysis of the vehicle's braking process is essential for determining a reasonable and effective warning duration. The detailed stages of the vehicle braking process are depicted in Fig. 7.

The vehicle braking process can be mathematically modeled as follows:

$$D_{veh} = \begin{cases} \frac{1}{36}\left(\tau_1 + \frac{\tau_2}{2}\right) v_{veh} + \frac{v_{veh}^2}{25.92 a_{veh}} v_{veh\_end} = 0 \\ \frac{1}{36}\left(\tau_1 + \frac{\tau_2}{2}\right) v_{veh} + \frac{v_{veh}^2 - v_{veh\_end}^2}{25.92 a_{veh}} v_{veh\_end} \neq 0 \end{cases} \quad (10)$$

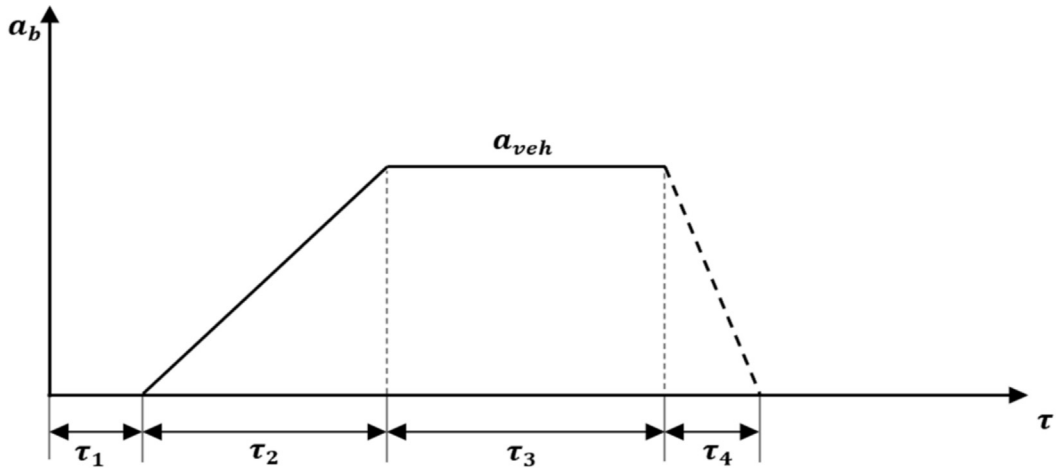

Fig. 7. Chronology of the Vehicle Braking Process.

where v_{veh} is the speed of the self-driving vehicle, a_{veh} represents the target braking deceleration, v_{veh_end} is the target speed of the vehicle at the final moment, τ_1 denotes the braking system adjustment time, and τ_2 corresponds to the time required for the braking deceleration to reach its full effect.

Fig. 8 presents the diagram of the vehicle's minimum safety distance. D_{safe} represents the relative distance between the vehicle and the obstacle ahead at the conclusion of active collision avoidance. L_{veh} indicates the braking distance of the vehicle, while L_{obs} refers to the distance to the obstacle ahead. The vehicle's safety distance,L, is expressed as:

$$L = L_{veh} - L_{obs} + D_{safe} \tag{11}$$

The vehicle safety distance model presented in this paper is designed to differentiate between three critical vehicle operations: safe driving, forward collision warning (FCW), and emergency braking. The model facilitates the calculation of essential safety distances, including the warning safety distance, the braking initiation safety distance, and the minimum braking safety distance for each operational scenario.

Assuming the obstacle ahead maintains a consistent motion state, the vehicle braking process is analyzed under three distinct scenarios: stationary, uniform motion or accelerated motion, and emergency braking of the obstacle. The safety distance calculations for these scenarios incorporate the varying motion states of the obstacle. This flexibility ensures the model's suitability for a wide range of real-world driving conditions. The relative distance at the end of the braking moment D_{safe}, is expressed as:

$$D_{safe} = \begin{cases} 3.6v_{veh} = 0 \\ max(0.2364v_{veh} + 1.6109, 3.6)v_{veh} > 0 \end{cases} \tag{12}$$

Where v_{veh} is the self-vehicle speed.

The trigger braking deceleration a_{veh_min} and the maximum braking deceleration a_{veh_max} are given by:

$$\begin{cases} a_{veh\_min} = min(\frac{4m}{s^2}, \mu g) \\ a_{veh\_max} = max(\frac{7m}{s^2}, \mu g) \end{cases} \tag{13}$$

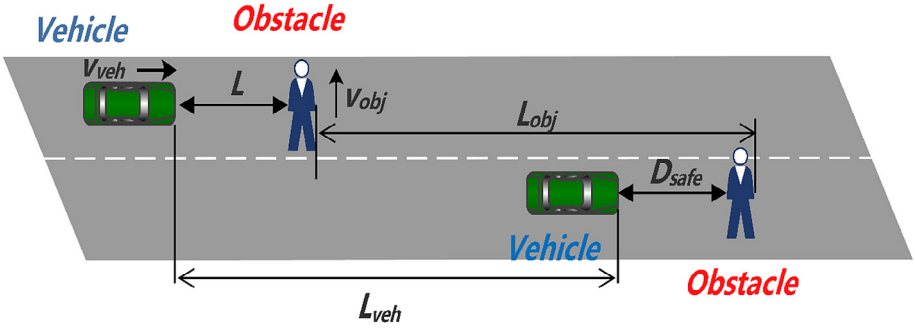

Fig. 8. Safety Distance Model for Vehicle Collision Avoidance.

where μ represents the coefficient of friction.

A summary of the safety distance calculation derivations for each case is provided below. The statuses of obstacles ahead are categorized as follows:

- Warning Safety Distance (L_w): Represents the critical distance at which a warning is issued to alert the driver of a potential collision.
- Braking Initiation Safety Distance (L_b): Denotes the distance at which braking should commence to ensure safe deceleration.
- Minimum Braking Safety Distance (L_s): Indicates the absolute minimum distance required to avoid a collision during braking.

(1) Stationary Obstacles

$$\begin{cases} \frac{1}{36}\left(\tau_1+\frac{\tau_2}{2}\right)v_{veh}+\frac{v_{veh}^2}{25.92a_{veh\_min}}+D_{safe}+\frac{1}{36}v_{veh}t_{dr} \\ \frac{1}{36}\left(\tau_1+\frac{\tau_2}{2}\right)v_{veh}+\frac{v_{veh}^2}{25.92a_{veh\_min}}+D_{safe} \\ \frac{1}{36}\left(\tau_1+\frac{\tau_2}{2}\right)v_{veh}+\frac{v_{veh}^2}{25.92a_{veh\_min}}+D_{safe} \end{cases} \tag{14}$$

(2) Obstacles in Uniform or Accelerated Motion

$$\begin{cases} \frac{1}{36}\left(\tau_1+\frac{\tau_2}{2}\right)\left(v_{veh}-\left(v_{obj}\right)+\frac{v_{veh}^2-v_{obj}^2}{25.92a_{veh\_min}}+D_{safe}+\frac{1}{36}v_{veh}t_{dr}\frac{1}{36}\left(\tau_1+\frac{\tau_2}{2}\right)\right. \\ \left(v_{veh}-\left(v_{obj}\right)+\frac{v_{veh}^2-v_{obj}^2}{25.92a_{veh\_min}}+D_{safe}\frac{1}{36}\left(\tau_1+\frac{\tau_2}{2}\right)\right)\left(v_{veh}-\left(v_{obj}\right)+\frac{v_{veh}^2-v_{obj}^2}{25.92a_{veh\_min}}+D_{safe}\right. \end{cases} \tag{15}$$

(3) Obstacles Undergoing Emergency Braking

$$\begin{cases} \frac{1}{36}\tau_1 v_{veh}+\frac{1}{7.2}\tau_2\left(v_{veh}-v_{obj}\right)+\frac{v_{veh}^2-v_{obj}^2}{25.92a_{veh\_min}}+D_{safe}+\frac{1}{36}v_{veh}t_{dr}\frac{1}{36}\tau_1 v_{veh}+\frac{1}{7.2}\tau_2\left(v_{veh}-v_{obj}\right) \\ +\frac{v_{veh}^2-v_{obj}^2}{25.92a_{veh\_min}}+D_{safe}\frac{1}{36}\tau_1 v_{veh}+\frac{1}{7.2}\tau_2\left(v_{veh}-v_{obj}\right)+\frac{v_{veh}^2-v_{obj}^2}{25.92a_{veh\_min}}+D_{safe} \end{cases} \tag{16}$$

In our research on pedestrian collision avoidance, the dynamic characteristics of pedestrian movement demand accurate modeling to ensure precise evaluation of collision risks. The position of a pedestrian changes over time, potentially leading to constraints that fail to meet real-world collision avoidance requirements. To address this, we transform the pedestrian's position at the actual moment of collision risk into a corresponding position constraint, refining the initial position constraint. This process accounts for various pedestrian movement scenarios, as illustrated in Fig. 9.

Based on the aforementioned front obstacle movement scenarios, determining the moment of collision risk requires analyzing the relative motion state between the vehicle and the obstacle along the lane's direction. The results of these calculations are summarized in the subsequent section. To ensure both safety and comfort during the emergency active collision avoidance process, this study introduces cost functions that evaluate critical parameters, including the vehicle's lateral speed, lateral acceleration, and lateral jerk.

$$J_y=\sum_{t=1}^{t_{end}}\left(p_t v_{y_t}^2+q_t a_{y_t}^2+r_t j_{y_t}^2\right) \tag{17}$$

Here, p_t, q_t, and r_t represent the weight values corresponding to the lateral velocity, lateral acceleration, and lateral jerk of the collision avoidance path, respectively. The coefficient of friction (μ) used in the safety distance equations is a critical factor in braking performance estimation. In our experimental setup, we use a typical value for dry asphalt (μ = 0.7), which is consistent with values reported in transportation safety standards. For real-world applications, μ can be dynamically estimated using onboard weather sensors, such as rain detectors, road surface cameras, or tire slip estimators, allowing for real-time adjustment under wet or icy conditions.

The cost function weights used in the collision risk evaluation (p_t, q_t, r_t) are empirically tuned to balance responsiveness, trajectory smoothness, and lateral safety. Specifically, p_t emphasizes time-to-collision minimization, p_t penalizes sharp steering deviations, and r_t controls the weight of lateral drift away from pedestrian paths. These values were calibrated using multiple driving sequences with varying pedestrian behaviors to reflect a practical balance between safety sensitivity and false positive suppression.

By assigning appropriate non-negative weight values, the smoothness and practicality of the collision avoidance path are effectively ensured.

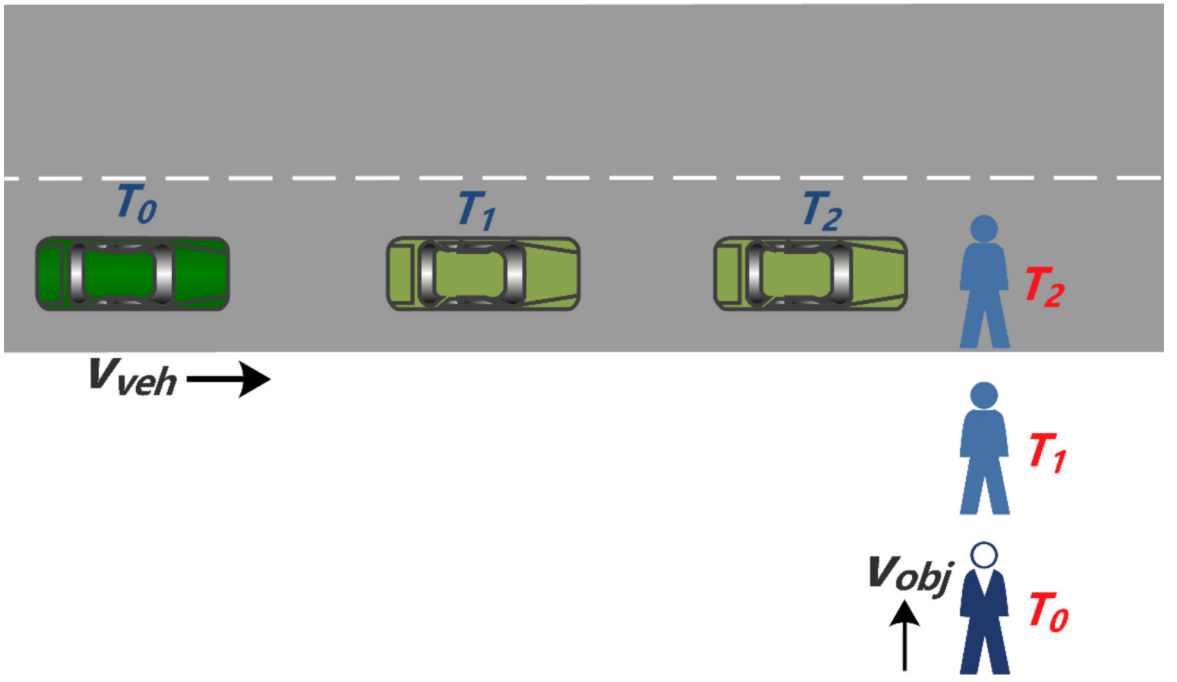

Fig. 9. Obstacle lateral drive-in position constraint schematic.

$$y_{min_t} \le y_t \le y_{max_t}, \forall t \in [1, N_{end}]$$

$$v_{min_t} \le v_{yt} \le v_{max_t}, \forall t \in [1, N_{end}]$$

$$a_{min_t} \le a_{yt} \le a_{max_t}, \forall t \in [1, N_{end}]$$

$$j_{min_t} \le j_{yt} \le j_{max_t}, \forall t \in [1, N_{end}]$$

where y_t, v_{yt}, $a_{yt} and j_{yt}$ represent the lateral position, lateral velocity, lateral acceleration, and lateral jerk of the self-vehicle at the sampling moment t, respectively. The lateral position y_t must comply with the vehicle's positional constraints, ensuring that the path always remains within the travelable area. Additionally, v_{yt}, $a_{yt} and j_{yt}$ must adhere to the vehicle's kinematic constraints. These parameters are influenced by the vehicle's dynamic capabilities, travel constraints, and the road adhesion coefficient, ensuring safe and stable operation during collision avoidance maneuvers.

The calculation of the collision hazard moment for pedestrians is influenced by the relative motion between the vehicle and the pedestrian, as well as the pedestrian's movement characteristics. The scenarios are as follows:

- **Pedestrian with High Longitudinal Acceleration ($a_{ped} \ge 0$)**

  The collision hazard moment T_{coll} is calculated as:

$$T_{coll} = \frac{L}{v_{veh} - v_{ped}} \tag{18}$$

  Here, v_{ped} is the pedestrian's velocity.

- **Pedestrian with Low Longitudinal Acceleration ($a_{ped} < 0$)**

o If the vehicle's braking trajectory satisfies: $v_{veh} T_{brake} > x_{brake} + L$

  In this scenario, the collision hazard moment T_{coll} is determined as:

$$T_{coll} = \frac{{v_{veh}}^2 + 2a_{veh}L}{2a_{veh}v_{veh}} \tag{19}$$

where: v_{veh} is the velocity of the vehicle, T_{brake} is the braking time, T_{brake} is the braking distance, L is the initial separation distance,

o If $v_{veh}T_{brake} \leq x_{brake} + L$, the calculation adjusts to account for shorter braking distances.

Alternatively, when the pedestrian's movement dynamics are significant:

$$T_{coll} = \frac{{v_{ped}}^2 - {v_{veh}}^2}{2a_{ped}\left(v_{veh} - v_{ped}\right)} + L \tag{20}$$

where $2a_{veh}$ is the longitudinal acceleration of the vehicle.

This model offers a versatile framework for assessing collision hazard moments in pedestrian detection scenarios, effectively addressing challenges posed by varying pedestrian movement patterns. It facilitates real-time decision-making in intelligent surveillance and autonomous driving systems, enhancing pedestrian safety.

4. Experimental results

4.1. Experimental setup

The experiments were conducted in a meticulously configured environment operating on the Windows 10 platform. The hardware setup comprised an Intel Core i7 processor with a 2.8 GHz clock speed, 16 GB of RAM, and an NVIDIA GeForce MX450 GPU. Model training was performed on Google Colab, leveraging the computational power of a Tesla A100 GPU (Tesla A100-SXM2-16 GB) with 15,109 MB of available memory. The experiments utilized Python 3.10.12 and PyTorch 2.1.0 as the core frameworks, ensuring robust compatibility and performance for deep learning tasks. To ensure consistency and reproducibility, uniform hyperparameters were utilized across all experiments, as outlined in Table 2.

The experiments were conducted with an input resolution of 640 × 640 pixels, a learning rate of 0.001, and a momentum parameter set to 0.937. The batch size was fixed at 64 for each training iteration. To mitigate overfitting, a weight decay of 0.005 was applied. Training spans 160 epochs, providing adequate time for model convergence while allowing the learning rate to decay progressively, optimizing performance. These meticulously controlled settings established a solid foundation for conducting fair and accurate comparative analyses across all tests.

4.2. Dataset description

In this section, we present two prominent datasets widely used in pedestrian detection research: the CityPersons dataset (Zhang et al., 2017) and the KITTI dataset (Geiger et al., 2012). The preprocessing phase is critical for enhancing these datasets, ensuring effective model training and analysis.

The KITTI dataset is a well-established benchmark for autonomous driving and object detection, providing a comprehensive collection of images captured from vehicles traversing diverse environments, including urban streets, rural areas, and highways in Germany. This data set encompasses a variety of challenging scenarios, such as objects with varying sizes, levels of occlusion, densities, and lighting conditions. Each image may include as many as 15 vehicles and 30 pedestrians, presenting substantial challenges for detection algorithms. The KITTI dataset includes eight object categories: car, van, truck, pedestrian, person sitting, and cyclist. For this study, we focus on the training set, comprising 7,481 labeled images, as the test set lacks ground truth labels. To enable a thorough evaluation, the training set is divided into three subsets using an 8:1:1 ratio, yielding 5,984 images for training, 748 for validation, and 749 for testing. This partition ensures a balanced assessment of the model's performance under diverse conditions.

The CityPersons dataset, derived from the Cityscapes dataset, is specifically tailored for pedestrian detection and includes detailed annotations. The dataset contains 2,975 training images, 500 validation images, and 1,575 testing images. Each image typically features an average of seven pedestrians, with annotations capturing both full-body and visible regions. Collected across 18 cities and spanning three seasons, the dataset encompasses a wide range of weather conditions, adding complexity to pedestrian detection tasks. With a total of 19,654 labeled pedestrian instances in high-resolution images (1024 × 2048 pixels), it provides a rich resource for model training. The validation set includes 500 images from three cities, ensuring diverse urban environments for evaluating model accuracy and robustness.

Table 2
Experimental Hyperparameter Details.

Parameter	Value
Learning Rate	0.001
Image Size	640 × 640
Momentum	0.937
Optimizer	SGD
Batch Size	32
Epochs	100
Weight Decay	0.005

4.3. Evaluation metrics

To assess the model's performance, we employed precision, recall, and mean average precision (mAP) as key evaluation metrics. Precision is the percentage of true positive predictions among all instances predicted as positive, whereas recall is the percentage of true positive predictions among all actual positive instances. The formulas for calculating precision and recall are:

$$Precision = TP/FP + TP \tag{21}$$

$$Recall = TP/TP + FN \tag{22}$$

Here, TP (True Positive) refers to correctly predicted positive samples, FP (False Positive) denotes incorrectly predicted positive samples, and FN (False Negative) indicates missed positive samples. Average Precision (AP) reflects the model's performance across varying precision-recall thresholds, calculated as:

$$AP = \int_0^1 p(r)dr \tag{23}$$

The mean Average Precision (mAP) metric provides an overall measure of the model's accuracy by averaging the AP values across all categories. This is expressed as:

$$mAP = \sum AP/N_{classes} \tag{24}$$

In the above equation, N_classes represent the total number, and APi is the average precision value of the i-th category. mAP@0.5 represents the average accuracy value. The number 0.5 represents the IoU (Intersection over Union) threshold of 0.5. In target detection, IoU represents the degree of overlap between the detected and actual target bounding boxes.

In this formula, *Nclasses* denotes the total number of object categories, while *APi* represents the average precision for the i-th category. The term mAP@0.5mAP@0.5mAP@0.5 refers to the mean average precision at a 0.5 Intersection over Union (IoU) threshold, which is used to measure the overlap between the predicted and ground truth bounding boxes. An IoU threshold of 0.5 indicates that the predicted bounding box must overlap with the ground truth by at least 50 % for the detection to be considered correct.

4.4. Ablation study

In this section, an ablation study is conducted to evaluate the individual contributions of each proposed module to the overall performance of the model. This involves systematically incorporating or excluding specific components during the training process and comparing the resulting performance metrics. By isolating each module's effect, its impact on detection accuracy can be accurately assessed.

The analysis focuses on three key components: the SimAM attention module, the GSConv convolutional blocks, and an additional 160×160 small-object detection head. Each of these components was incrementally integrated into the baseline YOLOv9 model, and the system's performance was evaluated using standard metrics, including Precision (P), Recall (R), mean Average Precision at IoU threshold 0.5 (mAP@0.5), and mean Average Precision across multiple thresholds (mAP@0.5:0.95). The developed models were trained and tested using two widely recognized benchmark datasets: CityPersons and KITTI. The results are summarized in Tables 3 and 4, presenting performance comparisons before and after incorporating the proposed modules.

The ablation results on the KITTI dataset demonstrate that each proposed module contributes incrementally to performance enhancement. Incorporating the SimAM attention module led to moderate improvements in all metrics, reflecting better spatial attention. Adding GSConv blocks further enhanced precision, recall, and mAP, suggesting improved feature representation. Finally, integrating the 160×160 detection head resulted in the highest performance across all metrics Precision (98.79 %), Recall (96.62 %), and mAP@0.5 (96.21 %) indicating a substantial gain in small-object detection capability and overall system robustness.

The ablation study on the CityPersons dataset confirms the consistent effectiveness of the proposed modules. Introducing the SimAM attention module yielded a noticeable boost in detection precision and recall, indicating enhanced focus on relevant pedestrian features. The inclusion of GSConv blocks further improved the model's capability to extract discriminative

Table 3
Ablation Study Results on the KITTI Dataset.

Model Variant	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Baseline YOLOv9	95.42	93.12	91.58	71.24
+ SimAM Attention	96.21	94.05	92.87	73.46
+ GSConv Integration	97.03	94.78	94.12	75.88
Final Model	98.79	96.62	96.21	78.35

Table 4
Ablation Study Results on the CityPersons Dataset.

Model Variant	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Baseline YOLOv9	93.87	91.24	89.73	69.42
+ SimAM Attention	94.66	92.58	91.02	71.33
+ GSConv Integration	95.52	93.10	92.56	73.11
Final Model	97.41	95.35	95.17	76.94

features, while the 160 × 160 detection head led to significant gains in detecting small-scale pedestrians. The final model achieved the best results, with Precision (97.85 %), Recall (95.43 %), and mAP@0.5 (95.07 %), demonstrating superior performance in dense urban environments.

Overall, the results confirm that each module contributes incrementally to the system's accuracy, and their combination leads to a substantial improvement in detection performance under real-world driving scenarios.

4.5. Data analysis

The proposed approach was evaluated using two well-established pedestrian detection datasets: the KITTI dataset and the CityPersons dataset. The results demonstrate a significant improvement in model performance as the number of training epochs increases. Table 5 presents the evaluation metrics for pedestrian detection on these datasets, including Precision, Recall, mAP@0.5 (mean Average Precision at a 50 % Intersection-over-Union threshold), and mAP@0.5:0.95 (mean Average Precision averaged across IoU thresholds ranging from 0.5 to 0.95). These metrics are reported for training durations of 10, 50, and 100 epochs, providing a comprehensive assessment of the model's performance across different training durations.

The results reveal consistent improvements across all performance metrics as training progresses. For the KITTI dataset, at 10 epochs, the model achieves a precision of 91.50 %, a recall of 82.25 %, an mAP@0.5 of 84.20 %, and an mAP@0.5:0.95 of 72.80 %. These initial results indicate effective pedestrian detection with potential for further enhancement. After 50 epochs, precision increases to 93.80 %, recall rises to 86.70 %, and mAP@0.5 improves to 92.10 %. By 100 epochs, precision reaches 95.25 %, recall improves to 91.85 %, and mAP@0.5 increases to 94.55 %, demonstrating significant gains in performance.

For the CityPersons dataset, the model starts with a precision of 92.40 % and a recall of 83.65 % at 10 epochs, outperforming the KITTI dataset at the same stage. The mAP@0.5 is 85.10 %, and mAP@0.5:0.95 is 75.35 %, reflecting strong initial performance. After 50 epochs, precision improves to 94.20 %, recall rises to 87.20 %, and mAP@0.5 reaches 93.00 %. By 100 epochs, precision increases to 96.10 %, recall improves to 92.75 %, and mAP@0.5 reaches 95.80 %, highlighting the model's strong progress in pedestrian detection.

These results demonstrate significant improvements in performance metrics as training epochs increase for both datasets. The steady rise in precision, recall, and mAP metrics underscores the effectiveness of extended training in enhancing pedestrian detection capabilities. This progress highlights the model's potential for real-world applications, particularly in autonomous driving and intelligent surveillance systems, where accurate and reliable pedestrian detection is critical.

The addition of the 160 × 160 detection head and the multi-scale fusion design led to a measurable improvement in small-pedestrian detection performance. On the CityPersons dataset, the model achieved a 4.2 % increase in average precision (AP) for objects under 50 pixels tall, compared to the baseline configuration. This confirms the value of incorporating FPN-inspired feature integration in addressing the scale sensitivity inherent to pedestrian-aware systems.

Fig. 10 illustrates the improvement in the proposed system's performance across training epochs. During the initial phases of training, both precision and recall exhibited rapid increases, highlighting the model's ability to adapt quickly to the task. As training progressed, these metrics continued to rise steadily, albeit at a slower pace, ultimately reaching substantial levels by the end of 100 epochs. Both metrics showed significant growth during the early stages of training, with mAP@0.5 consistently achieving higher values compared to mAP@0.5:0.95. As training advanced, performance stabilized, followed by gradual improvement. By the conclusion of training at 100 epochs, both metrics reached their peak values, demonstrating the model's remarkable ability to perform accurately and robustly.

This performance trajectory underscores the effectiveness of extended training and validates the model's capacity to adapt to complex scenarios, providing robust and reliable results for real-world applications.

Table 5
Performance Evaluation Metrics for the Developed System on the KITTI and CityPersons Datasets.

	KITTI dataset				CityPersons dataset			
	Precision	Recall	mAP50	mAP50-95	Precision	Recall	mAP50	mAP50-95
10	91.50	82.25	84.20	72.80	92.40	83.65	85.10	75.35
50	93.80	86.70	92.10	81.90	94.20	87.20	93.00	81.15
100	95.25	91.85	94.55	84.30	96.10	92.75	95.80	84.85

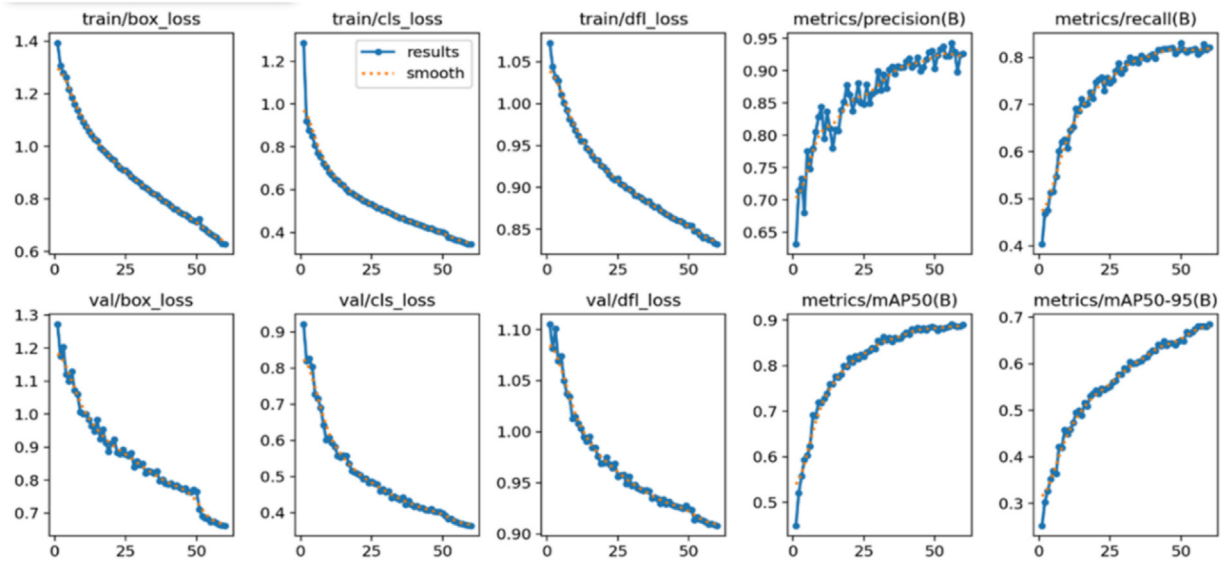

Fig. 10. Performance Results of the Proposed System.

In this research, testing images were captured using a vehicle's driving recorder in real traffic scenarios under various weather conditions. The visualized testing results demonstrate that the proposed system accurately detects most objects in each image, including those located at significant distances from the camera. This highlights the system's robustness and effectiveness across diverse environments.

Figs. 11 and 12 present qualitative results, showcasing both successful and failed detections, providing valuable insights into the system's detection capabilities. These figures illustrate the system's impressive progress, demonstrating its ability to recognize pedestrians with high precision and consistency. Even under challenging conditions, such as blurry backgrounds and varying object sizes, the model maintained reliable performance, further validating its applicability in real-world scenar-

Fig. 11. Qualitative Analysis of Pedestrian Detection in Diverse Scenarios: (a) Urban Streets, (b) Nighttime, (c) Pedestrians Near Motorcycles, (d) Pedestrians Near Special Vehicles.

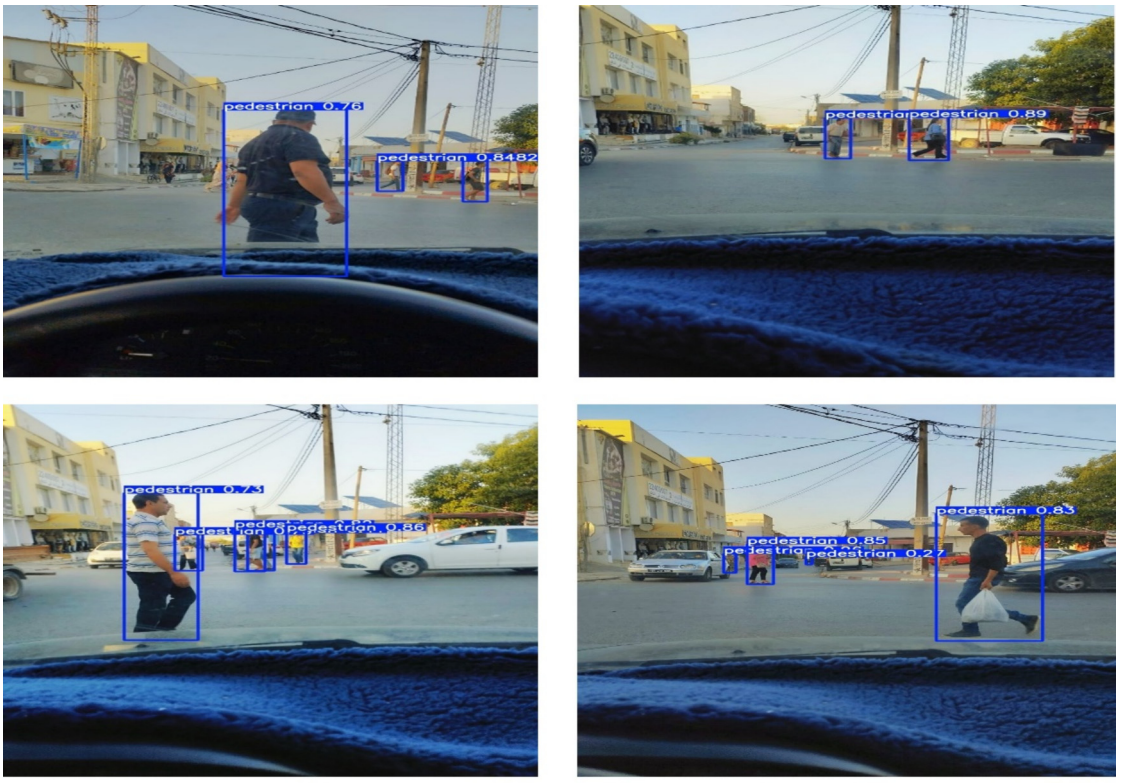

Fig. 12. Visualized successful testing cases of the whole system in various traffic scenes under different environment conditions.

ios. Overall, these results emphasize the system's potential for deployment in complex and dynamic environments, where accurate and consistent pedestrian detection is critical.

Table 6 presents the performance and processing time of the pedestrian detection tasks, emphasizing the time required for object detection with an input size of 640 × 640 pixels. This section evaluates the system's overall performance, particularly in terms of processing speed. The object detection task requires just 0.016 s per image with the specified input size. This processing time highlights the system's capability to meet real-time operational requirements, making it well-suited for applications that demand rapid decision-making, such as autonomous vehicles and intelligent surveillance systems. Although the system exhibits high accuracy, its real-time performance evidenced by notably low processing time further underscores its effectiveness in managing pedestrian detection tasks within dynamic and complex environments. This combination of accuracy and speed underpins its potential for practical deployment in scenarios where both are critical.

4.6. Accident warning validation

To rigorously assess the effectiveness of the proposed accident warning functionality, a series of dedicated experiments were conducted to evaluate the system's ability to issue timely alerts prior to potential pedestrian collisions. The experiments were designed to simulate real-world pedestrian scenarios, including jaywalking, partial occlusion, and varying speeds of approach. The warning system was assessed using four key metrics: warning precision, warning recall, average lead time before a collision (based on Time-to-Collision, TTC), and the successful early warning rate. Results were calculated over a test set comprising diverse urban sequences with dynamic pedestrian behavior.

As summarized in Table 7, the system achieved a warning precision of 97.6 %, a recall of 95.2 %, and an average lead time of 2.8 s before a potential collision, providing sufficient time for effective braking intervention. Furthermore, the early warning success rate reached 96.5 %, confirming the system's robustness in dynamic urban scenarios. These results demonstrate

Table 6
Performance and processing time of pedestrian detection tasks.

Task	Time (s)	Input size
Object detection	0.016	640 * 640

Table 7
Accident Warning System Performance Metrics.

Metric	Value
Warning Precision	97.6 %
Warning Recall	95.2 %
Average Lead Time Before Collision (seconds)	2.8
Percentage of Successful Early Warnings	96.5 %

that the accident warning module reliably delivers accurate and early alerts, thereby significantly enhancing the proactive safety capabilities of the proposed pedestrian-aware collision prevention system.

4.7. Comparison experiments

This section offers a comprehensive comparison of the performance of the YOLOv8n and the proposed YOLOv9 models on two widely recognized pedestrian detection datasets: the KITTI dataset and the CityPersons dataset. The evaluation emphasizes critical metrics, including Precision**,** Recall**,** and mAP at Intersection-over-Union (IoU) thresholds of 0.5 (mAP@0.5) and 0.5–0.95 (mAP@0.5:0.95).The results, summarized in Table 8, highlight the significant improvements achieved by the YOLOv9 model over the YOLOv8n model. These enhancements underscore the superior performance of YOLOv9 in terms of detection accuracy and robustness, reinforcing its suitability for advanced pedestrian detection tasks in diverse scenarios.

The comparison between the YOLOv8n and YOLOv9 models, as detailed in Table 8, underscores the substantial performance gains achieved with YOLOv9. On the KITTI dataset, YOLOv9 demonstrates notable advancements across all metrics. Precision increases by 2.41 %, reaching 98.79 %, signifying more accurate identification of pedestrians and a reduction in false positives. Recall improves by 3.53 %, rising from 88.65 % with YOLOv8n to 92.18 %, reflecting the enhanced capability of YOLOv9 to detect a higher proportion of true positives. In terms of mean Average Precision (mAP), YOLOv9 achieves an mAP@0.5 of 96.21 %, marking a 4.04 % improvement, and an mAP@0.5–0.95 of 84.32 %, a 4.93 % increase, indicating robust detection accuracy across a range of IoU thresholds. On the CityPersons dataset, YOLOv9 continues to outperform YOLOv8n across all metrics. It achieves a precision of 96.62 %, an improvement of 1.65 %, demonstrating greater accuracy in detecting pedestrians in urban settings. Recall increases by 4.67 %, reaching 90.81 %, showcasing its ability to capture more true positives. Additionally, YOLOv9 shows notable improvements in mAP metrics, with an mAP@0.5 of 94.83 % and an mAP@0.5–0.95 of 81.70 %, surpassing YOLOv8n by 3.27 % and 4.38 %, respectively.

These results clearly highlight YOLOv9′s consistent superiority over YOLOv8n on both datasets, achieving higher precision, recall, and mAP values. The architectural enhancements in YOLOv9 have effectively elevated its robustness and reliability for pedestrian detection in diverse and complex environments. These advancements position YOLOv9 as an excellent choice for applications such as autonomous driving and intelligent surveillance, where precise and efficient pedestrian detection is paramount.

The experimental results in Table 9 highlight the superior performance of the proposed pedestrian detection network relative to several benchmark models, including Faster R-CNN, SSD, YOLOv3, YOLOv5, YOLOv7, and YOLOv8, demonstrate varying detection capabilities. With a precision of 98.79 %, the proposed system sets a new benchmark in detection reliability. While SSD achieves high precision at 95.61 %, its low recall of 43.63 % indicates inconsistencies in detecting pedestrians under complex conditions. In contrast, our system maintains strong and consistent detection performance across varying scenes.

In terms of recall, the proposed system leads with 92.18 %, significantly outperforming Faster R-CNN at 72.93 % and SSD at 43.63 %, and surpassing modern models such as YOLOv7 and YOLOv8, both at 85.10 %. This demonstrates its strength in capturing true positives and maintaining robustness in crowded or occluded pedestrian scenarios. For average precision, the model records a mAP@0.5 of 96.21 % and mAP@0.5–0.95 of 84.32 %, which notably exceeds YOLOv8′s 92.30 % and 76.42 %, and YOLOv7′s 92.10 % and 76.75 %. These results highlight the model's ability to perform accurately across a range of intersection-over-union (IoU) thresholds, ensuring consistent localization quality.

Moreover, the model exhibits strong computational efficiency with only 34.7 million parameters, 86.2 GFLOPs, and an inference time of 11.2 ms per image. It outperforms YOLOv8 (68.1 M, 228.0 GFLOPs, 18.5 ms) and YOLOv7 (36.2 M, 105.4 GFLOPs, 16.5 ms), offering faster processing while delivering higher accuracy. This validates the model's real-time deployment capability in applications such as autonomous vehicles and intelligent surveillance systems. In conclusion, the pro-

Table 8
Performance Comparison between YOLOv8n and YOLOv9 Models on KITTI and CityPersons Datasets.

Model	KITTI dataset				CityPersons dataset			
	Precision	Recall	mAP50	mAP50-95	Precision	Recall	mAP50	mAP50-95
YOLOv8n	96.38	88.65	92.17	79.39	94.97	86.14	91.56	77.32
YOLOv9	98.79	92.18	96.21	84.32	96.62	90.81	94.83	81.70

Table 9
Accuracy and Computational Efficiency Comparison.

Models	Precision (%)	Recall (%)	mAP50	mAP50-95	Params (M)	FLOPs (G)	Inference Time (ms)
Faster R-CNN	39.94	72.93	65.14	51.23	52.2	180.4	89.7
SSD	95.61	43.63	68.14	52.36	34.6	70.1	33.2
YOLOV3	93.80	82.20	92.00	72.84	61.9	155.2	45.3
YOLOV5	93.80	82.00	90.93	75.15	45.3	96.2	24.1
YOLOV7	94.20	85.10	92.10	76.75	36.2	105.4	16.5
YOLOV8	94.90	85.10	92.30	76.42	68.1	228.0	18.5
Ours	98.79	92.18	96.21	84.32	34.7	86.2	11.2

posed system achieves an ideal trade-off between detection precision, robustness, and computational efficiency, positioning it as a state-of-the-art solution for real-time pedestrian detection in dynamic and safety–critical environments.

To approximate distributionally robust performance in the absence of formal optimization constraints, the proposed system was rigorously evaluated across a diverse set of real-world scenarios specifically designed to emulate distributional shifts. These scenarios included occluded pedestrian instances (e.g., partially hidden by vehicles or roadside structures), abrupt trajectory changes such as jaywalking or crossing from blind spots, significant illumination variation (including dusk, nighttime, and glare conditions), as well as visually complex urban environments with high background clutter and motion interference.

Despite these challenging and atypical conditions, the model consistently achieved high detection precision (above 95 %) and maintained reliable warning generation, with an average TTC lead time of 2.8 s. These results indicate strong resilience to out-of-distribution inputs and suggest the model's ability to generalize effectively beyond the conditions seen during training. While the study does not formalize distributional robustness through theoretical optimization (e.g., DRO frameworks), the systematic evaluation under adversarial and edge-case scenarios provides strong empirical support for the robustness and reliability of the proposed system under real-world operational variability.

5. Conclusion

In this paper, we introduce a novel system for pedestrian collision avoidance based on the YOLOv9 architecture. Extensive experiments conducted on the CityPersons and KITTI datasets highlight the superior performance of YOLOv9, achieving exceptional precision and mAP scores, significantly surpassing other mainstream networks in pedestrian detection. The proposed model demonstrates remarkable improvements in detecting small pedestrians, effectively addressing challenges such as small object detection, occlusions, tilted orientations, and complex backgrounds. It consistently recognizes pedestrians across diverse scenarios, including distant and occluded targets, as well as challenging environments.

Future work aims to expand the model's capabilities to cover a broader range of pedestrian categories, while optimizing its performance on various hardware and software platforms. Plans include the development of an embedded system leveraging FPGA and ARM processor boards for real-world testing in vehicular environments. Collaborations with local car manufacturers are also being pursued to enhance the practical applications of this research, with further advancements targeted at integrating pedestrian detection and tracking for autonomous driving systems and intelligent surveillance technologies.

CRediT authorship contribution statement

Wajdi Farhat: Writing – original draft, Validation, Software, Resources, Methodology, Conceptualization. **Marwa Geusani:** Software, Methodology. **Olfa Ben Rhaiem:** Writing – original draft, Visualization, Validation. **Radhia Zaghdoud:** Visualization, Validation, Supervision, Software. **Hassene Faiedh:** Supervision. **Chokri Souani:** Supervision.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available because they are part of an ongoing research project, and their public release could interfere with the study outcomes. However, they are available from the corresponding author upon reasonable request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alalwan, N., Abozeid, A., ElHabshy, A.A., Alzahrani, A., 2021. Efficient 3D deep learning model for medical image semantic segmentation. Alex. Eng. J. 60 (1), 1231–1239. https://doi.org/10.1016/j.aej.2020.10.046.

Arora, N., Kumar, Y., Karkra, R., Kumar, M., 2022. Automatic vehicle detection system in different environment conditions using fast R-CNN. Multimed. Tools Appl. 81 (13), 18715–18735. https://doi.org/10.1007/s11042-022-12347-8.

Asha, J., Giridhran, R., Agalya, K., Sathya, R., 2022. Traffic sign detection using HOG and GLCM with decision tree and random forest. In: 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS). Springer, Berlin, pp. 879–883. https://doi.org/10.1109/ICACRS55517.2022.10029118.

Assefa, A.A., Tian, W., Acheampong, K.N., Aftab, M.U., Ahmad, M., 2022. Small-scale and occluded pedestrian detection using multi mapping feature extraction function and modified soft-NMS. Comput. Intell. Neurosci. 2022, (1)9325803. https://doi.org/10.1155/2022/9325803.

Bin Zuraimi, M.A., Kamaru Zaman, F.H., 2021. Vehicle detection and tracking using YOLO and DeepSORT. In: 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), pp. 23–29. https://doi.org/10.1109/ISCAIE51753.2021.9431784.

Bouguettaya, A., Zarzour, H., Kechida, A., Taberkit, A.M., 2022. Vehicle detection from UAV imagery with deep learning: a review. IEEE Trans. Neural Netw. Learn. Syst. 33 (11), 6047–6067. https://doi.org/10.1109/TNNLS.2021.3080276.

Brunetti, A., Buongiorno, D., Trotta, G.F., Bevilacqua, V., 2018. Computer vision and deep learning techniques for pedestrian detection and tracking: a survey. Neurocomputing 300, 17–33. https://doi.org/10.1016/j.neucom.2018.01.092.

Chaurasiya, R., Ganotra, D., 2023. Deep dilated CNN based image denoising. Int. J. Inf. Technol. 15 (1), 137–148. https://doi.org/10.1007/s41870-022-01125-2.

Chen, L. et al, 2021. Deep neural network based vehicle and pedestrian detection for autonomous driving: a Survey. IEEE Trans. Intell. Transp. Syst. 22 (6), 3234–3246. https://doi.org/10.1109/TITS.2020.2993926.

Chen, X. et al, 2025. Intelligent ship route planning via an A∗ search model enhanced double-deep Q-network. Ocean Eng. 327, 120956. https://doi.org/10.1016/j.oceaneng.2025.120956.

Chen, X., Wu, H., Han, B., Liu, W., Montewka, J., Liu, R.W., 2023. Orientation-aware ship detection via a rotation feature decoupling supported deep learning approach. Eng. Appl. Artif. Intel. 125, 106686. https://doi.org/10.1016/j.engappai.2023.106686.

Costea, A.D., Varga, R., Nedevschi, S., 2017. Fast boosting based detection using scale invariant multimodal multiresolution filtered features. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 993–1002. https://doi.org/10.1109/CVPR.2017.112.

Dai, X. et al, 2021. Multi-task faster R-CNN for nighttime pedestrian detection and distance estimation. Infrared Phys. Technol. 115, 103694. https://doi.org/10.1016/j.infrared.2021.103694.

Dong, X., Xie, K., Yang, H., 2022. How did COVID-19 impact driving behaviors and crash severity? A multigroup structural equation modeling. Accid. Anal. Prev. 172, 106687. https://doi.org/10.1016/j.aap.2022.106687.

Eskandari Torbaghan, M., Sasidharan, M., Reardon, L., Muchanga-Hvelplund, L.C.W., 2022. Understanding the potential of emerging digital technologies for improving road safety. Accid. Anal. Prev. 166, 106543. https://doi.org/10.1016/j.aap.2021.106543.

Fang, J., Wang, F., Xue, J., Chua, T.-S., 2024. Behavioral intention prediction in driving scenes: a survey. IEEE Trans. Intell. Transp. Syst. 25 (8), 8334–8355. https://doi.org/10.1109/TITS.2024.3374342.

Fang, S., Zhang, B., Hu, J., 2023. improved mask R-CNN multi-target detection and segmentation for autonomous driving in complex scenes. Sensors 23, (8)8. https://doi.org/10.3390/s23083853.

Gao, J., Yu, B., Chen, Y., Bao, S., Gao, K., Zhang, L., 2024. An ADAS with better driver satisfaction under rear-end near-crash scenarios: a spatio-temporal graph transformer-based prediction framework of evasive behavior and collision risk. Transp. Res. Part C Emerg. Technol. 159, 104491. https://doi.org/10.1016/j.trc.2024.104491.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361. doi: 10.1109/CVPR.2012.6248074.

Gorrini, A., Crociani, L., Vizzari, G., Bandini, S., 2018. Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation. Transp. Res. Part F Traffic Psychol. Behav. 59, 269–285. https://doi.org/10.1016/j.trf.2018.09.016.

Gu, R., Li, Y., Cen, X., 2023. Exploring the stimulative effect on following drivers in a consecutive lane change using microscopic vehicle trajectory data. Transp. Saf. Environ. 5, (2)tdac047. https://doi.org/10.1093/tse/tdac047.

Guerrieri, M., Parla, G., 2022. Smart tramway systems for smart cities: a deep learning application in ADAS systems. Int. J. Intell. Transp. Syst. Res. 20 (3), 745–758. https://doi.org/10.1007/s13177-022-00322-4.

Ha, S.V.-U., Chung, N.M., Phan, H.N., Nguyen, C.T., 2020. TensorMoG: a tensor-driven Gaussian mixture model with dynamic scene adaptation for background modelling. Sensors 20, (23)23. https://doi.org/10.3390/s20236973.

Han, C., Gao, G., Zhang, Y., 2019. Real-time small traffic sign detection with revised faster-RCNN. Multimed. Tools Appl. 78 (10), 13263–13278. https://doi.org/10.1007/s11042-018-6428-0.

He, S., Yuan, Y., Yin, B., 2025. LMD_YOLO: a lightweight and efficient model for pavement defects detection. IEEE Access 13, 65510–65525. https://doi.org/10.1109/ACCESS.2025.3557938.

Hsu, W.-Y., Lin, W.-Y., 2021. Ratio-and-scale-aware YOLO for pedestrian detection. IEEE Trans. Image Process. 30, 934–947. https://doi.org/10.1109/TIP.2020.3039574.

Hussain, M., 2024. YOLOv1 to v8: unveiling each variant–a comprehensive review of YOLO. IEEE Access 12, 42816–42833. https://doi.org/10.1109/ACCESS.2024.3378568.

Ji, Y. et al, 2024. Toward autonomous vehicles: a survey on cooperative vehicle-infrastructure system. iScience 27, (5)109751. https://doi.org/10.1016/j.isci.2024.109751.

Kušić, K., Schumann, R., Ivanjko, E., 2023. A digital twin in transportation: real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics. Adv. Eng. Inf. 55, 101858. https://doi.org/10.1016/j.aei.2022.101858.

Labi, S., Chen, S., Dong, J., Li, Y., Sabu, J., John, A.P., 2024. Using virtual reality techniques to investigate interactions between fully autonomous vehicles and vulnerable road users. Center for Connected and Automated Transportation (CCAT). Technical Report. doi: 10.7302/22481.

Liang, X., Meng, X., Zheng, L., 2021. Investigating conflict behaviours and characteristics in shared space for pedestrians, conventional bicycles and e-bikes. Accid. Anal. Prev. 158, 106167. https://doi.org/10.1016/j.aap.2021.106167.

Liang, C., Zhang, Z., Zhou, X., Li, B., Zhu, S., Hu, W., 2022. Rethinking the competition between detection and ReID in multiobject tracking. IEEE Trans. Image Process. 31, 3182–3196. https://doi.org/10.1109/TIP.2022.3165376.

Losada, Á., Páez, F.J., Luque, F., Piovano, L., 2023. Effectiveness of the autonomous braking and evasive steering system OPREVU-AES in simulated vehicle-to-pedestrian collisions. Vehicles 5 (4), 1553–1569. https://doi.org/10.3390/vehicles5040084.

Muhammad, K., Ullah, A., Lloret, J., Ser, J.D., de Albuquerque, V.H.C., 2021. Deep learning for safe autonomous driving: current challenges and future directions. IEEE Trans. Intell. Transp. Syst. 22 (7), 4316–4336. https://doi.org/10.1109/TITS.2020.3032227.

Nafakh, A.J. et al, 2023. "Translation of driver-pedestrian behavioral models at semi-controlled crosswalks into a quantitative framework for practical self-driving vehicle applications. Part B (Pedestrian Volume Analytics) 61B. https://doi.org/10.5703/1288284317718.

Nascimento, A.M. et al, 2020. A systematic literature review about the impact of artificial intelligence on autonomous vehicle safety. IEEE Trans. Intell. Transp. Syst. 21 (12), 4928–4946. https://doi.org/10.1109/TITS.2019.2949915.

NHTSA Estimates for 2022 Show Roadway Fatalities Remain Flat After Two Years of Dramatic Increases | NHTSA. Accessed: Nov. 18, 2024. [Online]. Available: https://www.nhtsa.gov/press-releases/traffic-crash-death-estimates-2022.

Park, J., Kim, D., Huh, K., 2021. Emergency collision avoidance by steering in critical situations. Int. J. Automot. Technol. 22 (1), 173–184. https://doi.org/10.1007/s12239-021-0018-2.

Qiu, L., Zhang, D., Tian, Y., Al-Nabhan, N., 2021. Deep learning-based algorithm for vehicle detection in intelligent transportation systems. J. Supercomput. 77 (10), 11083–11098. https://doi.org/10.1007/s11227-021-03712-9.

Qu, J., Li, Y., Li, H., Liu, S., 2023. SimAM-based optimization algorithm for small target detection. In: Third International Conference on Advanced Algorithms and Neural Networks (AANN 2023). SPIE. pp. 170–175. doi: 10.1117/12.3005083.

Rezwana, S., Lownes, N., 2024. Interactions and behaviors of pedestrians with autonomous vehicles: a synthesis. Future Transp. 4 (3), 3. https://doi.org/10.3390/futuretransp4030034.

Sarkar, A., Hickman, J.S., McDonald, A.D., Huang, W., Vogelpohl, T., Markkula, G., 2021. Steering or braking avoidance response in SHRP2 rear-end crashes and near-crashes: a decision tree approach. Accid. Anal. Prev. 154, 106055. https://doi.org/10.1016/j.aap.2021.106055.

Shawky, M., Alsobky, A., Al Sobky, A., Hassan, A., 2023. Traffic safety assessment for roundabout intersections using drone photography and conflict technique. Ain Shams Eng. J. 14, (6)102115. https://doi.org/10.1016/j.asej.2023.102115.

Sormoli, A., Dianati, M., Mozaffari, S., Woodman, R., 2024. Optical flow based detection and tracking of moving objects for autonomous vehicles. IEEE Trans. Intell. Transp. Syst. 25 (9), 12578–12590. https://doi.org/10.1109/TITS.2024.3382495.

Tishby, N., Pereira, F.C., Bialek, W., 2000. The information bottleneck method, Apr. 24. arXiv: arXiv:physics/0004057. doi: 10.48550/arXiv.physics/0004057.

Wang, M., Ding, Y., Liu, Y., Qin, Y., Li, R., Tang, Z., 2025. MixSSC: forward-backward mixture for vision-based 3D semantic scene completion. IEEE Trans. Circuits Syst. Video Technol., 1 https://doi.org/10.1109/TCSVT.2025.3527235.

Wang, Z., Zhan, J., Duan, C., Guan, X., Lu, P., Yang, K., 2023. A review of vehicle detection techniques for intelligent vehicles. IEEE Trans. Neural Netw. Learn. Syst. 34 (8), 3811–3831. https://doi.org/10.1109/TNNLS.2021.3128968.

Wisesa, B.A., Wathan, M.H., Faristasari, E., Duli, S.A.J., Agustin, S., Swengky, B., 2025. Vehicle theft detection using YOLO based on license plates and vehicle ownership. Int. J. Inform. Comput. 7 (1), 73–85. https://doi.org/10.35842/ijicom.v7i1.105.

Yan, C., Zhang, H., Li, X., Yuan, D., 2022. R-SSD: refined single shot multibox detector for pedestrian detection. Appl. Intell. 52 (9), 10430–10447. https://doi.org/10.1007/s10489-021-02798-1.

Yang, L., Zhang, R.-Y., Li, L., Xie, X., 2021. SimAM: a simple, parameter-free attention module for convolutional neural networks. In: Proceedings of the 38th International Conference on Machine Learning, PMLR. pp. 11863–11874. Accessed: May 02, 2025. [Online]. Available: https://proceedings.mlr.press/v139/yang21o.html.

Yelchuri, R., Dash, J.K., Singh, P., Mahapatro, A., Panigrahi, S., 2022. Exploiting deep and hand-crafted features for texture image retrieval using class membership. Pattern Recogn. Lett. 160, 163–171. https://doi.org/10.1016/j.patrec.2022.06.017.

Zhang, C., Berger, C., 2023. Pedestrian behavior prediction using deep learning methods for urban scenarios: a review. IEEE Trans. Intell. Transp. Syst. 24 (10), 10279–10301. https://doi.org/10.1109/TITS.2023.3281393.

Zhang, S., Benenson, R., Schiele, B., 2017. CityPersons: a diverse dataset for pedestrian detection, Feb. 19. arXiv: arXiv:1702.05693. doi: 10.48550/arXiv.1702.05693.

Zhang, X. et al., 2024. LDConv: linear deformable convolution for improving convolutional neural networks, Jul. 22. arXiv: arXiv:2311.11587. doi: 10.48550/arXiv.2311.11587.

Zhang, Y., Qiao, Y., Fricker, J.D., 2020. Investigating pedestrian waiting time at semi-controlled crossing locations: application of multi-state models for recurrent events analysis. Accid. Anal. Prev. 137, 105437. https://doi.org/10.1016/j.aap.2020.105437.